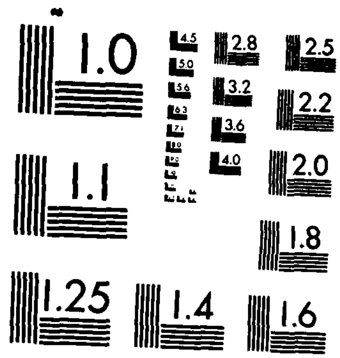
AD-A164 093 THE RELATIVE ENTROPY OF A RANDOM VECTOR WITH RESPECT TO 1/1
ANOTHER RANDOM VE..(U) PITTSBURGH UNIV PA CENTER FOR
MULTIVARIATE ANALYSIS M ROSENBLATT-ROTH OCT 85
UNCLASSIFIED TR-85-35 AFOSR-TR-86-0036 F49620-82-K-0001 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

THE RELATIVE ENTROPY OF A RANDOM VECTOR
WITH RESPECT TO ANOTHER RANDOM VECTOR

by

M. Rosenblatt-Roth[1]

Center for Multivariate Analysis

University of Pittsburgh

THE RELATIVE ENTROPY OF A RANDOM VECTOR
WITH RESPECT TO ANOTHER RANDOM VECTOR

by

M. Rosenblatt-Roth[1]

October 1985

Technical Report No. 85-35

Center for Multivariate Analysis
University of Pittsburgh
Fifth Floor, Thackeray Hall
Pittsburgh, PA 15260

[1]This work is sponsored by the Air Force Office of Scientific Research under Contract F49620-82-K-0001. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The concept of information in probability theory and statistics

Part A: The concept of relative entropy

Chapter I. The relative entropy of a random vector with respect to another random vector

I.0. Introduction

I.0.1. This is the first chapter of the first part of an extended work dedicated to the presentation of the author's original results, as yet unpublished, except [14], concerning the concept of information in its most general aspects.

This work is the development of the ideas presented by the author in his talk [15] at the Sixth International Symposium on Multivariate Analysis, July 25-29, 1983, organized by the Center for Multivariate Analysis, University of Pittsburgh. It was partially sponsored by the Center and the author expresses his sincere appreciation for this to Professor P. R. Krishnaiah and Professor C. R. Rao.

Basically, this work can be regarded as a development and continuation of the work done in this direction by A. N. Kolmogorov, I. M. Gelfand, A. M. Yaglom, R. L. Dobrushin, M. S. Pinsker, G. Kallianpur, between 1956-1960 as well as a continuation of the author's work between 1956-1982.

I.0.2. This chapter has an introductory character, covering some of the necessary preliminaries for the following chapters. It is divided in four subchapters. The first subchapter discusses the problems connected with the definition of the concept of relative entropy and the second some elementary properties of this concept; the third presents some additivity theorems while the fourth presents a generalization of this concept.

At variance with the following chapters, this chapter is presenting a number of results obtained by other authors. Because these results are spread in various publications, some difficult to find, in various languages, and a reference book does not exist, they are presented here.

It is to remark that some of the original proofs of those results are exceedingly difficult to follow, some representing only indications how the proofs should go, some are complicated without any reason, some contain non-necessary restrictions. Some results are given completely without any proof.

For all those reasons the author is presenting here complete straightforward proofs for all the results discussed; sometimes the results are presented with better proofs, sometime with new proofs, sometime the results themselves are bettered. The comments at the enc of the chapter will indicate the author's part in the proof or in bettering the proof if it is the case.

The third and fourth subchapters contain mainly results belonging to the author, some presented without proofs in [12], [13], or completely new ones.

I.0.3. Harold Jeffreys , professor of astronomy at the University of Cambridge, England, introduced in literature the concept of relative entropy. In a paper [5] presented for publication in 1974 he defines the quantity which in our notation is

$$h(\xi:\eta) + h(\eta:\xi)$$

as a measure of discrepancy between the probability distributions of the random variables ξ,η. In the second edition of his "Theory of probability" [6], he continues to discuss the properties and uses of this quantity. It is to remark that he did not name in any way this concept.

Beginning in 1951, S. Kullback started a sustained research effort, together with various associates, to solve a series of statistical problems with the help of this concept. [9][10]

Claude Shannon introduced in 1948 the concept of entropy of a random variable and the concept of the quantity of information contained in one random variable about another one, as basic concepts of information theory [16], [17].

It is easy to recognize that the concept of relative entropy and the concept of quantity of information can be obtained as particular cases of the concept of relative entropy.

From these inter-relations, Jeffrey's concept took its name but Jeffrey's name was forgotten inbetween.

I.1. The definition of relative entropy.

I.1.1. Let ξ, η be two random vectors, with the same values x_i, and let

$$P_\xi(x_i) = P(\xi = x_i),\ P_\eta(x_i) = P(\eta = x_i) \quad (1 \le i \le n). \qquad \text{(I.1.1.1)}$$

The relative entropy of ξ with respect to η, or of P_ξ with respect to P_η, is given by the expression

$$h(\xi : \eta) = h(P_\xi : P_\eta) = \sum_{i=1}^{n} P_\xi(x_i)\log\frac{P_\xi(x_i)}{P_\eta(x_i)} \qquad \text{(I.1.1.2)}$$

where for $a \ge 0$ we consider $0\log\frac{0}{a} = 0$.

I.1.2. Let now (Ω, Σ, P) be a probability space, where Ω is a set of elements ω, Σ a α- algebra of subsets of Ω, P a probability measure on Σ.

We consider two random vectors ξ, η, defined on this probability space, with values in the measure space (X, S, μ), where X is a set of elements x, S a σ - algebra of subsets of X, μ a measure on S. Let

$$P_\xi(T) = P\{\omega;\ \xi(\omega) \in T\},\ P_\eta(T) = P\{\omega;\ \eta(\omega) \in T\},\ T \in S \qquad \text{(I.1.2.1)}$$

be their probability measures.

Let $Z = \{Z_s\}$ be an S-measurable partition of X, i.e. a finite family of S-measurable non-overlapping sets Z_s, the union of which is X. Let us denote by U the set of all S-measurable partitions Z of X.

For any given random vectors ξ,η and for any S measurable partition Z of X, we define two finite valued random vectors ξ_Z, η_Z, both with the same values $s = 1, 2, \ldots, k$, where k is the number of elements in the partition Z, and such that

$$P_{\xi_Z}(s) = P_\xi(Z_s), \quad P_{\eta_Z}(s) = P_\eta(Z_s) \tag{I.1.2.2}$$

By (I.1.1.1), the relative entropy of ξ_Z with respect to η_Z is

$$h(\xi_Z : \eta_Z) = h(P_{\xi_Z} : P_{\eta_Z}) = \sum_{s=1}^{k} P_\xi(Z_s)\log\frac{P_\xi(Z_s)}{P_\eta(Z_s)} \tag{I.1.2.3}$$

I.1.3. Lemma I.1.

Let $P_i, Q_i > 0$ $(1 \le i \le n)$ and

$$P = \sum_{i=1}^{n} P_i, \quad Q = \sum_{i=1}^{n} Q_i \tag{I.1.3.1}$$

Then

$$P\log\frac{P}{Q} \le \sum_{i=1}^{n} P_i\log\frac{P_i}{Q_i} \tag{I.1.3.2}$$

with equality iff

$$P_i = Q_i \quad (1 \le i \le n) \tag{I.1.3.3}$$

Proof:

Let : $\phi(t)$ be a real valued continuous convex function, defined on the real line; $\alpha_i > 0$, t_i $(1 \le i \le n)$ arbitrary real numbers. By Jensen's inequality

$$\phi\left(\sum_{i=1}^{n} \alpha_i t_i\right) \le \sum_{i=1}^{n} \alpha_i \phi(t_i) \tag{I.1.3.4}$$

where equality takes place iff

$$t_1 = t_2 = \ldots = t_n \qquad \text{(I.1.3.5)}$$

Taking in (I.1.3.4)

$$\phi(t) = t\log t,\ \alpha_i = \frac{Q_i}{Q},\ t_i = \frac{P_i}{P}\ (1 \le i \le n) \qquad \text{(I.1.3.6)}$$

we obtain (I.1.3.2), and from (I.1.3.5) we obtain (I.1.3.3).

I.1.4. Let $Z, Z' \in U$. We say that Z' is a subpartition of Z if each element of Z can be represented as the union of some elements of Z'. The set U is partial ordered by this relation, which we denote by $Z' < Z$. Indeed

a) $Z' < Z$ and $Z < Z'$ imply that Z, Z' are identical.

b) $Z'' < Z'$ and $Z' < Z$ imply that $Z'' < Z$.

c) For any $Z', Z'' \in U$ it exists an element $Z \in U$ such that $Z < Z'$, $Z < Z''$.

Indeed if $Z' = \{Z'_{s'}\}$, $Z'' = \{Z''_{s''}\}$, we may take $Z = \{Z_{s',s''}\}$ with $Z_{s',s''} = Z'_{s'} \cap Z''_{s''}$.

I.1.5. <u>Lemma I.2</u>

<u>If</u> $Z, Z' \in U$ <u>and</u> $Z' < Z$, <u>then</u>

$$h(\xi_Z : \eta_Z) \le h(\xi_{Z'} : \eta_{Z'}) \qquad \text{(I.1.5.1)}$$

<u>Proof</u>: Suppose Z consists of elements $Z_s \in S$ $(1 \le s \le n)$ and Z' consists of elements $Z'_{s'} \in S$ $(1 \le s' \le n')$ $(n \le n')$.

Let

$$Z_s = \bigcup_{s' \in L_s} Z'_{s'} \qquad (1 \le s \le n) \qquad \text{(I.1.5.2)}$$

where L_s is some subset of $1, 2, \ldots, n'$, so that L_s, L_t are not overlapping if $s \ne t$, and the union of all L is $(1, 2, \ldots, n')$. Then

$$P_\xi(Z_s) = \sum_{s' \in L_s} P_\xi(Z'_{s'}),\quad P_\eta(Z_s) = \sum_{s' \in L_s} P_\eta(Z'_{s'}) \qquad \text{(I.1.5.3)}$$

From Lemma I.1 it follows that

$$\sum_{s'\varepsilon L_s} P_\xi(Z'_{s'})\log\frac{P_\xi(Z'_{s'})}{P_\eta(Z'_{s'})} \geq \left[\sum_{s'\varepsilon L_s} P_\xi(Z'_{s'})\right]\log\frac{\sum_{s'\varepsilon L_s} P_\xi(Z'_{s'})}{\sum_{s'\varepsilon L_s} P_\eta(Z'_{s'})}$$

$$= P_\xi(Z_s)\log\frac{P_\xi(Z_s)}{P_\eta(Z_s)} \quad (1 \leq s \leq n) \qquad (\text{I.1.5.4})$$

so that

$$\sum_{s=1}^{n}\sum_{s'\varepsilon L_s} P_\xi(Z'_{s'})\log\frac{P_\xi(Z'_{s'})}{P_\eta(Z'_{s'})} \geq \sum_{s=1}^{n} P_\xi(Z_s)\log\frac{P_\xi(Z_s)}{P_\eta(Z_s)} \qquad (\text{I.1.5.5})$$

or

$$\sum_{s'=1}^{n'} P_\xi(Z'_{s'})\log\frac{P_\xi(Z'_{s'})}{P_\eta(Z'_{s'})} \geq \sum_{s=1}^{n} P_\xi(Z_s)\log\frac{P_\xi(Z_s)}{P_\eta(Z_s)} \qquad (\text{I.1.5.6})$$

i.e. (I.1.5.1)

Definition I.1. The quantity

$$h(\xi : \eta) = \sup_{Z\varepsilon U} h(\xi_z : \eta_Z) \qquad (\text{I.1.5.7})$$

is the relative entropy of ξ with respect to η, or of P_ξ with respect to P_η.

Theorem I.1. For arbitrary random vectors ξ, η

$$h(\xi : \eta) \geq 0 \qquad (\text{I.1.5.8})$$

with equality iff

$$P_\xi = P_\eta (\text{a.e.}) \qquad (\text{I.1.5.9})$$

Proof. If ξ, η are finite valued, let us consider the result in Lemma I.1. with $P_i = P_\xi(x_i)$, $Q_i = P_\eta(x_i)$, $P = Q = 1$. So from (I.1.3.2) it follows I.1.5.8 and from (I.1.3.3) it follows $P_\xi(x_i) = P_\eta(x_i)$ $(1 \leq i \leq n)$. From Definition I.1. follows the result in (I.1.5.8), (I.1.5.9) in general.

I.1.6. In what follows, we will need the following result, in which we denote

$$E \,\Delta\, E_0 = (E - E_0) \bigcup (E_0 - E).$$

Lemma I.3.

a) If λ_K $(1 \le k \le r)$ is a sequence of σ-finite measures on an algebra L which generates the σ-algebra S, then for any set $E \varepsilon S$ for which $\lambda_k(E) < \infty$ $(1 \le k \le r)$ and for any positive number $\varepsilon > 0$ there exists a set $E_0 \varepsilon L$ such that

$$\lambda_k(E \,\Delta\, E_0) \le \varepsilon \;(1 \le k \le r). \tag{I.1.6.1}$$

b) Moreover, if $E^{(j)}$ $(1 \le j \le m)$ are non-overlapping sets belonging to S, then

$$\lambda_k \left[\left(\bigcup_{j=1}^{m} E^{(j)} \right) \Delta \left(\bigcup_{j=1}^{m} E_0^{(j)} \right) \right] < m\varepsilon, \quad (1 \le k \le r) \tag{I.1.6.2}$$

Proof. We prove first the above lemma in the case $r = 1$, and we denote $\lambda_1 = \lambda$. The proof of this case is performed in two steps.

The first part of this proof repeats that of Theorem D pp. 56 in [4]. We reproduce it here for the necessity to use the intermediary results for the second part of our proof. Because $E \varepsilon S$, it follows

$$\lambda(E) = \inf\{ \sum_{i=1}^{\infty} \mu(E_i);\; E \subset \bigcup_{i=1}^{\infty} E_i,\; E_i \varepsilon L;\; i = 1, 2, \ldots\} \tag{I.1.6.3}$$

so that it follows that there exists a sequence $\{E_i\}$ of sets in L such that

$$E \subset \bigcup_{i=1}^{\infty} E_i \tag{I.1.6.4}$$

and

$$\lambda\left(\bigcup_{i=1}^{\infty} E_i\right) \le \lambda(E) + \frac{\varepsilon}{2} \tag{I.1.6.5}$$

Since

$$\lim_{n\to\infty} \lambda(\bigcup_{i=1}^{n} E_i) = \lambda(\bigcup_{i=1}^{\infty} E_i) \qquad (I.1.6.6)$$

there exists a positive integer N such that for any $n > N$, considering the set

$$E_0 = \bigcup_{i=1}^{n} E_i \qquad (I.1.6.7)$$

it follows that

$$\lambda(\bigcup_{i=1}^{\infty} E_i) \leq \lambda(E_0) + \frac{\varepsilon}{2} \qquad (I.1.6.8)$$

Obviously

$$E_0 \in L. \qquad (I.1.6.9)$$

Because

$$\lambda(E - E_0) \leq (\bigcup_{i=1}^{\infty} E_i - E_0) = \lambda(\bigcup_{i=1}^{\infty} E_i) - \lambda(E_0) \leq \frac{\varepsilon}{2} \qquad (I.1.6.10)$$

$$\lambda(E_0 - E) \leq \lambda(\bigcup_{i=1}^{n} E_i - E) \leq \lambda(\bigcup_{\varepsilon=1}^{\infty} E_i - E) =$$

$$= \lambda(\bigcup_{i=1}^{\infty} E_i) - \lambda(E) \leq \frac{\varepsilon}{2} \qquad (I.1.6.11)$$

it follows that

$$\lambda(E \,\Delta\, E_0) = \lambda(E - E_0) + \lambda(E_0 - E) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \qquad (I.1.6.12)$$

i.e. (I.1.6.1)

The second part of the proof uses these results. Indeed, let us consider m non-overlapping sets $E^{(j)} \in L$ $(1 \leq j \leq m)$. Then from (I.1.6.7), (I.1.6.8) it is possible to find for each $E^{(j)}$ a sequence $\{E_i^{(j)}\}$ of sets in L such that

$$E^{(j)} \subset \bigcup_{i=1}^{\infty} E_i^{(j)} \qquad (I.1.6.4')$$

and

$$\lambda\left(\bigcup_{i=1}^{\infty} E_i^{(j)}\right) \leq \lambda(E^{(j)}) + \frac{\varepsilon}{2} \qquad \text{(I.1.6.5')}$$

and obviously, because $E^{(j)} \cap E^{(k)} = \phi(j \neq k)$, we can find sets $E_i^{(j)}$ so that

$$E_e^{(j)} \cap E_r^{(k)} = \phi \ (j \neq k);\ e,\ r = 1,\ 2,\ \ldots;\ j,\ k = 1,\ 2,\ \ldots,\ m. \qquad \text{(I.1.6.13)}$$

But from (I.1.6.6) it follows

$$\lim_{n\to\infty} \lambda\left(\bigcup_{i=1}^{n} E_i^{(j)}\right) = \lambda\left(\bigcup_{i=1}^{\infty} E_i^{(j)}\right) / \ (1 \leq j \leq m) \qquad \text{(I.1.6.6')}$$

and so it exists a positive integer N_j such that denoting

$$E_0^{(j)} = \bigcup_{i=1}^{n} E_i^{(j)} \qquad \text{(I.1.6.7')}$$

for any $n > N_j$, then

$$\lambda\left(\bigcup_{i=1}^{\infty} E_i^{(j)}\right) \leq \lambda(E_0^{(j)}) + \frac{\varepsilon}{2}; \quad (1 \leq j \leq m) \qquad \text{(I.1.6.8')}$$

Let

$$N = \max_{1 \leq j \leq m} N_j \qquad \text{(I.1.6.14)}$$

From (I.1.6.7') it follows

$$\bigcup_{j=1}^{m} E_0^{(j)} = \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} E_i^{(j)} \qquad \text{(I.1.6.15)}$$

From (I.1.6.5') it follows that

$$\sum_{j=1}^{m} \lambda\left(\bigcup_{i=1}^{\infty} E_i^{(j)}\right) \leq \sum_{j=1}^{m} \lambda(E^{(j)}) + m\frac{\varepsilon}{2} \qquad \text{(I.1.6.5'')}$$

and from (I.1.6.13) this inequality can be written as

$$\lambda\left(\bigcup_{i=1}^{\infty} \bigcup_{j=1}^{m} E_i^{(j)}\right) \leq \lambda\left(\bigcup_{j=1}^{m} E^{(j)}\right) + m\frac{\varepsilon}{2} \qquad \text{(I.1.6.5''')}$$

From (I.1.6.8') it follows that

$$\sum_{j=1}^{m} \lambda\left(\bigcup_{i=1}^{\infty} E_i^{(j)}\right) \leq \sum_{j=1}^{m} \lambda(E_0^{(j)}) + \frac{\varepsilon}{2} \qquad \text{(I.1.6.8'')}$$

and from (I.1.6.13) this inequality can be written as

$$\lambda(\bigcup_{i=1}^{\infty} \bigcup_{j=1}^{m} E_i^{(j)}) \le \lambda(\bigcup_{j=1}^{m} E_0^{(j)}) + m\frac{\varepsilon}{2} \qquad \text{(I.1.6.8'')}$$

Obviously,

$$\bigcup_{j=1}^{m} E_0^{(j)} \in L \qquad \text{(I.1.6.16)}$$

From (I.1.6.8''') it follows that

$$\lambda(\bigcup_{j=1}^{m} E^{(j)} - \bigcup_{j=1}^{m} E_0^{(j)}) \le \lambda(\bigcup_{i=1}^{\infty} \bigcup_{j=1}^{m} E_i^{(j)} - \bigcup_{j=1}^{m} E_0^{(j)})$$

$$= \lambda(\bigcup_{i=1}^{\infty} \bigcup_{j=1}^{m} E_i^{(j)}) - \lambda(\bigcup_{j=1}^{m} E_0^{(j)}) \le m\frac{\varepsilon}{2} \qquad \text{(I.1.6.17)}$$

From (I.1.6.5''') it follows that

$$\lambda(\bigcup_{j=1}^{\infty} \bigcup_{j=1}^{m} E_i^{(j)} - \bigcup_{j=1}^{m} E^{(j)}) = \lambda(\bigcup_{i=1}^{\infty} \bigcup_{j=1}^{m} E_i^{(j)}) - \lambda(\bigcup_{j=1}^{m} E^{(j)}) \le m\frac{\varepsilon}{2} \qquad \text{(I.1.6.18)}$$

From (I.1.6.17), (I.1.6.18) it follows that

$$\lambda\left[(\bigcup_{j=1}^{m} E^{(j)}) \,\Delta\, (\bigcup_{j=1}^{m} E_0^{(j)})\right] \le m\varepsilon \qquad \text{(I.1.6.19)}$$

so that our lemma is proved for $r = 1$. In order to prove it for $r > 1$, let us consider in the above results

$$\lambda = \sum_{K=1}^{r} \lambda_r. \qquad \text{(I.1.6.20)}$$

From

$$\lambda_k(E) \le \lambda(E) \qquad \text{(I.1.6.21)}$$

it follows the general result stated in the Lemma I.3.

I.1.7. Lemma I.4.

Let μ be some finite measure on (X, S) and A, B $\in$ S. Then

$$|\mu(A) - \mu(B)| \leq \mu(A \Delta B) \qquad \text{(I.1.7.1)}$$

Proof. Because

$$\mu(A) = \mu(A \cap B) + \mu(A - B)$$

$$\mu(B) = \mu(A \cap B) + \mu(B - A)$$

it follows that

$$\mu(A) - \mu(B) = \mu(A - B) - \mu(B - A)$$

so that

$$|\mu(A) - \mu(B)| = |\mu(A - B) - \mu(B - A)|$$

$$\leq \mu(A - B) + \mu(B - A) = \mu(A \Delta B)$$

We say that an algebra L of S measurable sets generates the σ-algebra S_0 if S is the smallest σ-algebra such that $L \subset S$.

Theorem I.2.

Let:

a) L be an algebra of S-measurable sets, which generates S;

b) R be a family of S-measurable partitions of X.

If any partition consisting of sets from L has a subpartition belonging to R, then

$$h(\xi : \eta) = \sup_{Z \varepsilon R} h(\xi_Z : \eta_Z) \qquad \text{(I.1.7.2)}$$

Proof. Let $U_L \subset U$ be the totality of partitions Z of X, whose elements belong to the algebra L of S-measurable sets, and let

$$h_L(\xi : \eta) = \sup_{Z \varepsilon U_L} h(\xi_Z : \eta_Z) \qquad \text{(I.1.7.3)}$$

where $h(\xi_Z : \eta_Z)$ is given by (I.1.2.3). With this notation,

$$h_S(\xi : \eta) = h(\xi : \eta) \qquad \text{(I.1.7.4)}$$

Similarly, let

$$h_R(\xi : \eta) = \sup_{Z \varepsilon R} h(\xi_Z : \eta_Z) \qquad \text{(I.1.7.5)}$$

Let $Z = \{Z_s\}$ be a partition of X. From Lemma I.3, with $\lambda_1 = P_\xi$, $\lambda_2 = P_\eta$ it follows that for any $\varepsilon > 0$ and for all $Z_s \in S(1 \leq s \leq n)$ we can find non-overlapping sets $Z_{(0)s} \in L$ such that

$$P_\xi(Z_s \Delta Z_{(0)s}) \leq \varepsilon,\ P_\eta(Z_s \Delta Z_{(0)s}) \leq \varepsilon \quad (1 \leq s \leq n) \tag{I.1.7.6}$$

Let

$$Z_1^{(0)} = Z_{(0)1} \in L \tag{I.1.7.7}$$

$$Z_k^{(0)} = Z_{(0)k} - \bigcup_{i=1}^{k-1} Z_{(0)i} \in L \quad (1 \leq k \leq n) \tag{I.1.7.8}$$

$$Z_{n+1}^{(0)} = X - \bigcup_{i=1}^{n} Z_{(0)i} \in L \tag{I.1.7.9}$$

Because the family of sets $Z_s^{(0)}$ $(1 \leq s \leq n+1)$ is constituted of non-overlapping sets and their union is X, it follows that $Z^{(0)} = \{Z_s^{(0)}\}$ is a partition of X with $Z^{(0)} \in U_L$.

Using the results in Lemma I.4, it follows that for the measure P, which can be P_ξ or P_η, we have the inequalities

$$\begin{aligned} |P(Z_k^{(0)}) - P(Z_k)| &\leq |P(Z_k^{(0)}) - P(Z_{(0)k})| + |P(Z_{(0)k}) - P(Z_k)| \leq \\ &\leq P(Z_k^{(0)} \Delta Z_{(0)k}) + P(Z_{(0)k} \Delta Z_k) \quad (1 \leq k \leq n) \end{aligned} \tag{I.1.7.10}$$

Obviously,

$$\begin{aligned} Z_k^{(0)} \Delta Z_{(0)k} &= [Z_{(0)k} - \bigcup_{i=1}^{k-1} Z_{(0)i}] \Delta Z_{(0)k} \\ &= Z_{(0)k} - [Z_{(0)k} - \bigcup_{i=1}^{k-1} Z_{(0)i}] = (\bigcup_{i=1}^{k-1} Z_{(0)i}) \cap Z_{(0)k} \\ &\qquad (1 \leq k \leq n) \end{aligned} \tag{I.1.7.11}$$

We will show that

$$Z_k^{(0)} \Delta Z_{(0)k} \subset \bigcup_{i=1}^{k} (Z_i \Delta Z_{(0)i}) \quad (1 \leq k \leq n) \tag{I.1.7.12}$$

Indeed, let us decompose $Z_{(0)i}$ as

$$Z_{(0)i} = (Z_i \cap Z_{(0)i}) \cup (Z_{(0)i} - Z_i) \quad (1 \leq i \leq n) \tag{I.1.7.13}$$

So that

$$\bigcup_{i=1}^{k-1} Z_{(0)i} = \left[\bigcup_{i=1}^{k-1} (Z_i \cap Z_{(0)i})\right] \bigcup \left[\bigcup_{i=1}^{k-1} (Z_{(0)i} - Z_i)\right]$$

$$(i \leq k \leq n) \qquad \text{(I.1.7.14)}$$

$$Z_{(0)k} = (Z_k \cap Z_{(0)k}) \cup (Z_{(0)k} - Z_k) \quad (1 \leq k \leq n) \qquad \text{(I.1.7.15)}$$

From (I.1.7.11), (I.1.7.14), (I.1.7.15), because $Z_i (1 \leq i \leq n)$ are non-overlapping, it follows that

$$Z_k^{(0)} \Delta Z_{(0)k} = (\bigcup_{i=1}^{k-1} Z_{(0)i}) \cap Z_{(0)k} = \left[(\bigcup_{i=1}^{k-1} (Z_i - Z_{(0)i})) \cap (Z_{(0)k} - Z_k)\right] \bigcup$$

$$\subset \left[(\bigcup_{i=1}^{k-1} (Z_{(0)i} - Z_i)\right] \cup (Z_{(0)k} - Z_k) = \bigcup_{i=1}^{k} (Z_{(0)i} - Z_i) \subset \bigcup_{i=1}^{k} (Z_i \Delta Z_{(0)i}) \subset$$

which proves (I.1.7.12). From (I.1.7.12) it follows

$$P(Z_k^{(0)} \Delta Z_{(0)k}) \leq \sum_{i=1}^{k} P(Z_i \Delta Z_{(0)i}) \leq k\varepsilon \qquad \text{(I.1.7.16)}$$

so that from (I.1.7.10) we obtain the inequality

$$|P(Z_k^{(0)}) - P(Z_k)| \leq k\varepsilon + \varepsilon = (k+1)\varepsilon \leq (n+1)\varepsilon \quad (1 \leq k \leq n) \qquad \text{(I.1.7.17)}$$

so that

$$|P_\xi(Z_k^{(0)}) - P_\xi(Z_k)| \leq (n+1)\varepsilon \quad (1 \leq k \leq n) \qquad \text{(I.1.7.18)}$$

$$|P_\eta(Z_k^{(0)}) - P_\eta(Z_k)| \leq (n+1)\varepsilon \quad (1 \leq k \leq n) \qquad \text{(I.1.7.19)}$$

It follows that

$$P_\xi(Z_k^{(0)}) = P_\xi(Z_k) + \delta_{\xi,k}; \; |\delta_{\xi,k}| \leq (n+1)\varepsilon \qquad \text{(I.1.7.20)}$$

$$P_\eta(Z_k^{(0)}) = P_\eta(Z_k) + \delta_{\eta,k}; \; |\delta_{\eta,k}| \leq (n+1)\varepsilon \qquad \text{(I.1.7.21)}$$

and consequently

$$T_k = P_\xi(Z_k^{(0)}) \log \frac{P_\xi(Z_k^{(0)})}{P_\eta(Z_k^{(0)})} - P_\xi(Z_k) \log \frac{P_\xi(Z_k)}{P_\eta(Z_k)} = (P_\xi(Z_k) + \delta_{\xi,k}) \log \frac{P_\xi(Z_k) + \delta_{\xi,k}}{P_\eta(Z_k) + \delta_{\eta,k}} -$$

$$- P_\xi(Z_k) \log \frac{P_\xi(Z_k)}{P_\eta(Z_k)} = \left[P_\xi(Z_k) + \delta_{\xi,k}\right] \left[\log \frac{P_\xi(Z_k)}{P_\eta(Z_k)} + \log \frac{1 + \dfrac{\delta_{\xi,k}}{P_\xi(Z_k)}}{1 + \dfrac{\delta_{\eta,k}}{P_\eta(Z_k)}}\right]$$

$$- P_\xi(Z_k) \log \frac{P_\xi(Z_k)}{P_\eta(Z_k)} = (P_\xi(Z_k) + \delta_{\xi,k}) \log \frac{1+\varepsilon_{\xi,k}}{1+\varepsilon_{\eta,k}} + \delta_{\xi,k} \log \frac{P_\xi(Z_k)}{P_\eta(Z_k)}$$

$$\text{(I.1.7.22)}$$

where

$$\varepsilon_{\xi,k} = \frac{\delta_{\xi,k}}{P_\xi(Z_k)} \; ; \; \varepsilon_{\eta,k} = \frac{\delta_{\eta,k}}{P_\eta(Z_k)} \qquad \text{(I.1.7.23)}$$

So

$$T_k = P_\xi(Z_k^{(0)}) \log \frac{1+\varepsilon_{\xi,k}}{1+\varepsilon_{\eta,k}} + (n+1)\varepsilon \log \frac{P_\xi(Z_k)}{P_\eta(Z_k)} \qquad \text{(I.1.7.24)}$$

Let us denote

$$M = \min \{ \min_{1\le k\le n} P_\xi(Z_k); \; \min_{1\le k\le n} P_\eta(Z_k)\} \qquad \text{(I.1.7.25)}$$

Considering the case $h(\xi:\eta) < \infty$ it follows $M > 0$. Let us denote also

$$e = \frac{n+1}{M} \qquad \text{(I.1.7.26)}$$

Then

$$|\varepsilon_{\xi,k}| < \frac{|\delta_{\xi,k}|}{P_\xi(Z_k)} < \frac{(n+1)\varepsilon}{M} = e\varepsilon \qquad \text{(I.1.7.27)}$$

$$|\varepsilon_{\eta,k}| < \frac{|\delta_{\eta,k}|}{P_\eta(Z_k)} < \frac{(n+1)\varepsilon}{M} = e\varepsilon \qquad \text{(I.1.7.28)}$$

From $\log x < x-1$ it follows

$$\log \frac{1+\varepsilon_{\xi,k}}{1+\varepsilon_{\eta,k}} < \frac{1+\varepsilon_{\xi,k}}{1+\varepsilon_{\eta,k}} - 1 = \frac{\varepsilon_{\xi,k} - \varepsilon_{\eta,k}}{1+\varepsilon_{\eta,k}} \qquad \text{(I.1.7.29)}$$

from which it follows

$$\left|\log \frac{1+\varepsilon_{\xi,k}}{1+\varepsilon_{\eta,k}}\right| < \left|\frac{\varepsilon_{\xi,k} - \varepsilon_{\eta,k}}{1+\varepsilon_{\eta,k}}\right| < \frac{|\varepsilon_{\xi,k}| + |\varepsilon_{\eta,k}|}{1 - |\varepsilon_{\eta,k}|} \qquad \text{(I.1.7.30)}$$

and from (I.1.7.28) it follows

$$1 - |\varepsilon_{\eta,k}| > 1 - \varepsilon e \qquad \text{(I.1.7.30')}$$

so that with the help of (I.1.7.27), because $\varepsilon < 1$, it follows

$$\left|\log \frac{1+\varepsilon_{\xi,k}}{1+\varepsilon_{\eta,k}}\right| < \frac{2e\varepsilon}{1-e\varepsilon} \quad \frac{2e\varepsilon}{1-e} \qquad \text{(I.1.7.31)}$$

so that from (I.1.7.24) we obtain

$$|T_k| \le \left|\log \frac{1+\epsilon_{\xi,k}}{1+\epsilon_{\eta,k}}\right| + (n+1)\varepsilon \left|\log \frac{P_\xi(Z_k)}{P_\eta(Z_k)}\right| <$$

$$< \frac{2e\varepsilon}{1-e} + (n+1)\varepsilon \left|\log \frac{P_\xi(Z_k)}{P_\eta(Z_k)}\right| < \frac{2e\varepsilon}{1-e} + (n+1)\varepsilon \left|\log \frac{1}{M}\right| = a.e. \tag{I.1.7.32}$$

where

$$a = \frac{2e}{1-e} + (n+1)\varepsilon |\log M| \tag{I.1.7.32'}$$

Consequently,

$$\left|\sum_{k=1}^{n} T_k\right| \le \sum_{k=1}^{n} |T_k| < na\varepsilon \tag{I.1.7.33}$$

which can be written as

$$\left|\sum_{s=1}^{n} P_\xi(Z_s^{(0)}) \log \frac{P_\xi(Z_s^{(0)})}{P_\eta(Z_s^{(0)})} - \sum_{s=1}^{n} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)}\right| < na\varepsilon \tag{I.1.7.33'}$$

From

$$P\,[(\bigcup_{i=1}^{n} Z_i) \,\Delta\, (\bigcup_{i=1}^{n} Z_{(0)i})] < \varepsilon \tag{I.1.7.34}$$

because

$$\bigcup_{i=1}^{n} Z_i = X$$

it follows for $P = P_\xi$ or $P = P_\eta$, that

$$P(X \,\Delta \bigcup_{i=1}^{n} Z_{(0)i}) = P(X - \bigcup_{i=1}^{n} Z_{(0)i}) \le \varepsilon \tag{I.1.7.35}$$

or from (I.1.7.9)

$$P(Z_{n+1}^{(0)}) \le \varepsilon \tag{I.1.7.36}$$

i.e.

$$P_\xi(Z_{n+1}^{(0)}) \le \varepsilon, \quad P_\eta(Z_{n+1}^{(0)}) \le \varepsilon \tag{I.1.7.36'}$$

Let us denote

$$T_{n+1} = P_\xi(Z_{n+1}^{(0)}) \log \frac{P_\xi(Z_{n+1}^{(0)})}{P_\eta(Z_{n+1}^{(0)})} \tag{I.1.7.37}$$

so that from (I.1.7.27) and (I.1.7.36) it follows that

$$|T_{n+1}| = \left| P_\xi(Z^{(0)}_{n+1}) \log \frac{P_\xi(Z^{(0)}_{n+1})}{P_\eta(Z^{(0)}_{n+1})} \right| < \varepsilon \left|\log \frac{1}{M}\right| = \varepsilon |\log M| \tag{I.1.7.38}$$

Let

$$h(\xi_{Z^{(0)}} : \eta_{Z^{(0)}}) = \sum_{s=1}^{n+1} P_\xi(Z^{(0)}_s) \log \frac{P_\xi(Z^{(0)}_s)}{P_\eta(Z^{(0)}_s)} \tag{I.1.7.39}$$

$$h(\xi_Z : \eta_Z) = \sum_{s=1}^{n} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \tag{I.1.7.40}$$

so that

$$h(\xi_{Z^{(0)}} : \eta_{Z^{(0)}}) - h(\xi_Z : \eta_Z) = \sum_{k=1}^{n} T_k + T_{n+1} \tag{I.1.7.41}$$

so that from (I.1.7.33), (I.1.7.38) it follows

$$|h(\xi_{Z^{(0)}} : \eta_{Z^{(0)}}) - h(\xi_Z : \eta_Z)| \leq \left| \sum_{k=1}^{n} T_k \right| + |T_{n+1}| < (na + |\log M|)\varepsilon \tag{I.1.7.42}$$

Consequently, $h(\xi_{Z^{(0)}} : \eta_{Z^{(0)}})$ is as close as we want to $h(\xi_Z : \eta_Z)$, if the last one is finite, if we take ε sufficient small. If (I.1.7.40) is not finite, then (I.1.7.39) will be as large as we want, taking ε sufficient small.

Consequently, to any partition $Z \in U = U_s$, it corresponds a partition $Z^0 \in U_L$ such that (I.1.7.39), (I.1.7.40) are as close as we want, so that from the definitions of $h_S(\xi:\eta)$, $h_L(\xi:\eta)$ it follows that

$$h_S(\xi:\eta) \leq h_L(\xi:\eta) \tag{I.1.7.43}$$

From this, because any partition with elements in L has a subpartition in R, it follows from (I.1.7.3), (I.1.7.5), that

$$h_L(\xi:\eta) \leq h_R(\xi:\eta) \tag{I.1.7.44}$$

Because $U_L \subset U_S$, $U_R \subset U_S$ it follows

$$h_L(\xi:\eta) \leq h_S(\xi:\eta) \tag{I.1.7.45}$$

$$h_R(\xi:\eta) \leq h_S(\xi:\eta) \tag{I.1.7.46}$$

From (I.1.7.43), (I.1.7.44), (I.1.7.45), (I.1.7.46) it follows that

$$h_L(\xi:\eta) = h_R(\xi:\eta) = h_S(\xi:\eta) = h(\xi:\eta) \tag{I.1.7.47}$$

which proves theorem I.2.

Consequently, in the definition of $h(\xi:\eta)$ instead of considering all measurable partitions in U_S, we may consider only the subclass R. For example, in the case when X is the real line and S the σ-algebra of all Borel sets in it, it is sufficient to consider only the class R of partitions with elements in the algebra of finite unions of intervals. In the case when X is the cartezian product $X_1,\ldots,X_n$ of n real lines and S is the σ-algebra of all Borel sets in it, it is sufficient to consider only the class R of partitions with elements in the algebra of finite unions of n-dimensional intervals of the form $A_1 x \ldots x A_n$, where A_i is an interval on the real line ($1 \leq i \leq n$).

I.1.8. Theorem I.3

In order that the relative entropy of ξ with respect to η be finite, it is necessary that the probability distribution P_ξ be absolutely continuous with respect to the probability distribution P_η. Under this condition, the relative entropy $h(\xi:\eta)$ defined as the supremum (I.1.5.7) over all partitions of the range of ξ and η into a finite number of sets measurable with respect to P_ξ and P_η, is equal to the following integral

$$h(\xi:\eta) = \int_X a_{\xi:\eta}(x) \log a_{\xi:\eta}(x)\, P_\eta(dx) \tag{I.1.8.1}$$

where $a_{\xi:\eta}(x)$ is the Radon-Nicodym derivative of P_ξ with respect to P_η,

$$a_{\xi:\eta}(x) = \frac{P_\xi(dx)}{P_\eta(dx)}\,. \tag{I.1.8.2}$$

In this integral representation formula, the integral exists in the sense that the integral over the set where the integral is negative converges. In particular, $h(\xi:\eta)$ is finite or not according as this integral is finite or not.

Obviously, the integral representation formula can be written as

$$h(\xi:\eta) = \int_X i_{\xi:\eta}(x)\; P_\xi(dx) \tag{I.1.8.3}$$

where

$$i_{\xi:\eta}(x) = \log a_{\xi:\eta}(x) \tag{I.1.8.4}$$

is the relative entropy density of P_ξ with respect to P_η. In the particular case that the propability measures P_ξ, P_η are defined in terms of densities $\pi_\xi(x)$, $\pi_\eta(x)$ with respect to μ, the integral representation formula reduces to

$$h(\xi:\eta) = \int_X \pi_\xi(x)\; \log \frac{\pi_\xi(x)}{\pi_\eta(x)} \,.\, dx \tag{I.1.8.5}$$

where the integration is on μ-measure, and

$$a_{\xi:\eta}(x) = \frac{\pi_\xi(x)}{\pi_\eta(x)} \;. \tag{I.1.8.6}$$

Obviously, in the particular case that X is a countable space of elements x_i, and P_ξ, P_η are given as (I.1.1.1), then from (I.1.8.2) it follows

$$a_{\xi:\eta}(x_i) = \frac{P_\xi(x_i)}{P_\eta(x_i)} \tag{I.1.8.7}$$

and from (I.1.8.4) it follows

$$i_{\xi:\eta}(x_i) = \log \frac{P_\xi(x_i)}{P_\eta(x_i)} \tag{I.1.8.8}$$

and the integral representation formula reduces to (I.1.1.2) with $n = \infty$.

Proof of theorem I.3

First part of the proof. If P_ξ is not absolute continuous with respect to P_η, then there exists a set $B \in S$ such that $P_\xi(B) > 0$, $P_\eta(B) = 0$. Considering the partition $Z \in U$ consisting of the two elements $Z_1 = B$, $Z_2 = X - B$, it follows that

$$h(\xi_Z:\, \eta_Z) = \sum_{s=1}^{2} P_\xi(Z_s)\; \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \tag{I.1.8.9}$$

is not finite, and so is also

$$h(\xi:\eta) = \sup_{Z \in U} h(\xi_Z:\, \eta_Z) \tag{I.1.8.10}$$

<u>Second part of the proof</u>. Let us consider that P_ξ is absolutely continuous with respect to P_η.

In what follows, let P be some probability measure on the real line, such that

$$\int_0^\infty P(du) = 1 \qquad \text{(I.1.8.11)}$$

If $\phi(u)$ is some convex function on $[0,\infty)$, Jensen's inequality gives

$$\int_0^\infty \phi(u)\ P(du) \le \phi\left(\int_0^\infty u\ P(du)\right) \qquad \text{(I.1.8.12)}$$

Let

$$\phi(u) = -u \log u \qquad \text{(I.1.8.13)}$$

Then

$$\phi''(u) = -\frac{1}{u} < 0 \quad (0 \le u < \infty) \qquad \text{(I.1.8.14)}$$

so $\phi(u)$ is convex on $(0 \le u < \infty)$, and from (I.1.8.14) it follows the inequality

$$\int_0^\infty (u \log u)\ P(du) \ge \left[\int_0^\infty u\ P(du)\right] \log\left[\int_0^\infty u\ P(du)\right] \qquad \text{(I.1.8.15)}$$

Let $T \in S$, such that $P_\eta(T) > 0$. We define the measure P_T on S by the relation

$$P_T(A) = P_\eta\ \{[x;\ a_{\xi:\eta}(x) \in A] / x \in T\},\ A \in S \qquad \text{(I.1.8.16)}$$

where the bar means conditional probability. Let f(u) be some Borel measurable function on $[0,\infty)$. Then

$$\int_0^\infty f(u)\ P_T(du) = \int_0^\infty f(u) . \frac{1}{P_\eta(T)} . P_\eta\{[x;\ u < a_{\xi:\eta}(x) < u + du] \cap T\} =$$

$$= \frac{1}{P\ (T)} . \int_T f\ [a_{\xi:\eta}(x)] . P_\eta(dx) \qquad \text{(I.1.8.17)}$$

If $f(u) = 1$, from (I.1.8.17) it follows that

$$\int_0^\infty P_T(dx) = \frac{1}{P_\eta(T)} \cdot P_T(T) = 1 \qquad (I.1.8.18)$$

i.e. the measure P_T is a probability measure.

If $f(u) = u$, from (I.1.8.17) it follows that

$$\int_0^\infty u\, P_T(dx) = \frac{1}{P_\eta(T)} \cdot \int_T a_{\xi:\eta}(x).\, P_\eta(dx) = \frac{P_\xi(T)}{P_\eta(T)} \qquad (I.1.8.19)$$

If $f(u) = u \log u$, from (I.1.8.17) it follows that

$$\int_0^\infty u \log u.\ P_T(dx) = \frac{1}{P_\eta(T)} \cdot \int_T a_{\xi:\eta}(x) \log a_{\xi:\eta}(x).\ P_\xi(dx)$$

$$= \frac{1}{P_\eta(T)} \cdot \int_T [\log a_{\xi:\eta}(x)].\ P_\xi(dx) \qquad (I.1.8.20)$$

From (I.1.8.5), because of (I.1.8.19), (I.1.8.20) it follows

$$\frac{1}{P_\eta(T)} \cdot \int_T [\log a_{\xi:\eta}(x)] \cdot P_\xi(dx) \geq \frac{P_\xi(T)}{P_\eta(T)} \cdot \log \frac{P_\xi(T)}{P_\eta(T)}$$

or

$$P_\xi(T) \log \frac{P_\xi(T)}{P_\eta(T)} \leq \int_T [\log a_{\xi:\eta}(x)].\ P_\xi(dx) \qquad (I.1.8.21)$$

Now, let Z_s be elements of some partition Z of X. With $T = Z_s$, from (I.1.8.21) it follows

$$P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \leq \int_{Z_s} [\log a_{\xi:\eta}(x)].\ P_\xi(dx) \qquad (I.1.8.22)$$

and consequently

$$h(\xi_Z:\eta_Z) = \sum_s P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \leq \sum_s \int_{Z_s} [\log a_{\xi:\eta}(x)].\ P_\xi(dx) =$$

$$= \int_X [\log a_{\xi:\eta}(x)].\ P_\xi(dx) \qquad (I.1.8.23)$$

so from (I.1.5.7) it follows

$$h(\xi:\eta) \leq \int_X [\log a_{\xi:\eta}(x)].\ P_\xi(dx) \qquad (I.1.8.24)$$

Third part of the proof.

Because

$$\lim_{x\to 0}(x \log x) = 0, \lim_{k\to\infty} P_\xi\{x; |\log a_{\xi:\eta}(x)| > k\} = 0 \quad (I.1.8.25)$$

it follows that we may chose $\varepsilon > 0$ so small, and then take $k > 0$ so large that

$$-\frac{\varepsilon}{2} \leq P_\xi\{x; |\log a_{\xi:\eta}(x)| > k\} \log P_\xi\{x; |\log a_{\xi:\eta}(x)| > k\} \leq 0 \quad (I.1.8.26)$$

But

$$\{x; |\log a_{\xi:\eta}(x)| \leq k\} = \{x; e^{-k} \leq a_{\xi:\eta}(x) \leq e^k\} \quad (I.1.8.27)$$

Let Z_s $(1 \leq s \leq n)$ be such disjoint sets in S, that

a) $$\bigcup_{s=1}^{n} Z_s = \{x; |\log a_{\xi:\eta}(x)| \leq k\} \quad (I.1.8.28)$$

b) $$\log e_s - \log m_s \leq \frac{\varepsilon}{2} \quad (1 \leq s \leq n) \quad (I.1.8.28')$$

where

$$e_s = \sup\{a_{\xi:\eta}(x); x \in Z_s\}; m_s = \inf\{a_{\xi:\eta}(x); x \in Z_s\} \quad (I.1.8.29)$$

Let us define the set

$$Z_{n+1} = \{x; \log a_{\xi:\eta}(x)| > k\} \quad (I.1.8.30)$$

so that

$$\bigcup_{s=1}^{n+1} Z_s = X \quad (I.1.8.31)$$

and consequently Z_s $(1 \leq s \leq n+1)$ form a partition $Z^{(0)}$ of X.

Obviously for any s such that $1 \leq s \leq n$,

$$P_\xi(Z_s) = \int_{Z_s} a_{\xi:\eta}(x)\, P_\eta(dx) \leq e_s . \int_{Z_s} P_\eta(dx) = e_s . P_\eta(Z_s) \quad (I.1.8.32)$$

$$P_\xi(Z_s) = \int_{Z_s} a_{\xi:\eta}(x)\, P_\eta(dx) \geq m_x . \int_{Z_s} P_\eta(dx) = m_s . P_\eta(Z_s) \quad (I.1.8.32')$$

Similarly

$$\int_{Z_s} [\log a_{\xi:\eta}(x)].P_\xi(dx) \le \int_{Z_s} [\log e_s] P_\xi(dx) = (\log e_s).P_\xi(Z_s) \quad \text{(I.1.8.33)}$$

$$\int_{Z_s} [\log a_{\xi:\eta}(x)].P_\xi(dx) \ge \int_{Z_s} [\log m_s] P_\xi(dx) = (\log m_s).P_\xi(Z_s) \quad \text{(I.1.8.33')}$$

From (I.1.8.32), (I.1.8.32') it follows

$$m_s \le \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \le e_s \quad \text{(I.1.8.34)}$$

and from (I.1.8.33), (I.1.8.33') it follows

$$[\log m_s].P_\xi(Z_s) \le \int_{Z_s} [\log a_{\xi:\eta}(x)] P_\xi(dx) \le [\log e_s].P_\xi(Z_s) \quad \text{(I.1.8.35)}$$

From (I.1.8.22), (I.1.8.35) it follows

$$P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \le \int_{Z_s} [\log a_{\xi:\eta}(x)].P_\xi(dx) \le [\log e_s].P_\xi(Z_s) \quad \text{(I.1.8.36)}$$

so that from (I.1.8.35), (I.1.8.36) it follows

$$P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{Z_s} [\log a_{\xi:\eta}(x)].P_\xi(dx) \le [\log e_s].P_\xi(Z_s) - [\log m_s].P_\xi(Z_s) =$$

$$= (\log e_s - \log m_s).P_\xi(Z_s) \quad \text{(I.1.8.37)}$$

From (I.1.8.34), (I.1.8.36) it follows similarly

$$P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{Z_s} [\log a_{\xi:\eta}(x)].P_\xi(dx) \ge [\log e_s].P_\xi(Z_s) - [\log m_s].P_\xi(Z_s)$$

$$= (\log e_s - m_s).P_\xi(Z_s) \quad \text{(I.1.8.37')}$$

From (I.1.8.37), (I.1.8.37') it follows the inequality

$$\left| P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{Z_s} \log a_{\xi:\eta}(x).P_\xi(dx) \right| \le (\log e_s - \log m_s).P_\xi(Z_s); \; (1 \le s \le n)$$

$$\text{(I.1.8.38)}$$

from which it follows

$$\left| \sum_{s=1}^{n} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{\bigcup_{s=1}^{n} Z_s} \log a_{\xi:\eta}(x) . P_\xi(dx) \right| =$$

$$= \left| \sum_{s=1}^{n} [P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)}) - \int_{Z_s} \log a_{\xi:\eta}(x) . P_\xi(dx)] \right|$$

$$\leq \sum_{s=1}^{n} \left| P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{Z_s} \log a_{\xi:\eta}(x) . P_\xi(dx) \right| \leq$$

$$\leq \sum_{s=1}^{n} (\log e_s - \log m_s) P_\xi(Z_s) \leq \sum_{s=1}^{n} \frac{\varepsilon}{2} . P_\xi(Z_s) = \frac{\varepsilon}{2} . P(\bigcup_{s=1}^{n} Z_s) \leq \frac{\varepsilon}{2}$$

So

$$\left| \sum_{s=1}^{n} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{X - Z_{n+1}} \log a_{\xi:\eta}(x) . P_\xi(dx) \right| \leq \frac{\varepsilon}{2} \qquad \text{(I.1.8.39)}$$

From (I.1.8.26) we obtain

$$-\frac{\varepsilon}{2} \leq P_\xi(Z_{n+1}) \log P_\xi(Z_{n+1}) \leq P_\xi(Z_{n+1}) \log \frac{P_\xi(Z_{n+1})}{P_\eta(Z_{n+1})} \qquad \text{(I.1.8.40)}$$

From (I.1.8.39) it follows

$$-\frac{\varepsilon}{2} \leq \sum_{s=1}^{n} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{X-Z_{n+1}} \log a_{\xi:\eta}(x) . P_\xi(dx) \qquad \text{(I.1.8.41)}$$

From (I.1.8.40), (I.1.8.41) by addition, it follows

$$-\varepsilon \leq \sum_{s=1}^{n+1} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} - \int_{X-Z_{n+1}} \log a_{\xi:\eta}(x) . P_\xi(dx)$$

or

$$\sum_{s=1}^{n+1} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \geq \int_{X-Z_{n+1}} \log a_{\xi:\eta}(x) P_\xi(dx) - \varepsilon \qquad \text{(I.1.8.42)}$$

or

$$h(\xi_{Z(0)}; \eta_{Z(0)}) \geq \int_{X-Z_{n+1}} \log a_{\xi:\eta}(x) . P_\xi(dx) - \varepsilon \qquad \text{(I.1.8.43)}$$

from which it follows

$$h(\xi:\eta) = \sup_{Z\varepsilon U} h(\xi_Z:\eta_Z) \geq h(\xi_{Z^{(0)}}:\eta_{Z^{(0)}}) \geq [\log a_{\xi:\eta}(x)].P_\xi(dx) - \varepsilon$$

$$\geq \int_{\{x;|\log a_{\xi:\eta}(x)| \leq k\}} \qquad (I.1.8.44)$$

and if $k \to \infty$, $\varepsilon \to 0$, we obtain the inequality

$$h(\xi:\eta) \geq \int_X \log a_{\xi:\eta}(x).P_\xi(dx) \qquad (I.1.8.45)$$

From (I.1.8.24), (I.1.8.45) it follows (I.1.8.1).

If the integral (I.1.8.1), let us make the substitution

$$a_{\xi:\eta}(x) = u \qquad (I.1.8.46)$$

or

$$x = a^{-1}_{\xi:\eta}(u) \qquad (I.1.8.46')$$

This transforms the integrand in

$$u \log u \qquad (I.1.8.47)$$

and the measure $P_\eta(dx)$ is transformed in some measure $L(du)$. So, (I.1.8.1) takes the form

$$\int_{-\infty}^{\infty} (u \log u)L(du) \qquad (I.1.8.48)$$

Let us denote

$$f(u) = u \log u. \qquad (I.1.8.49)$$

Obviously

$$\{u;\ f(u) < 0\} = \{u;\ 0 \leq u < 1\} \qquad (I.1.8.50)$$

From

$$\frac{df}{du} = 1 + \log u \qquad (I.1.8.51)$$

it is seen that $f(u)$ has a minimum value for $u = e^{-1}$, and so

$$f(u) \geq f(e^{-1}) = -e^{-1} \qquad (I.1.8.52)$$

so that

$$\int_0^1 u \log u.L(du) \geq \int_0^1 -e^{-1}.L(du) = -e^{-1}.L([0,1]) \geq -e^{-1} \qquad (I.1.8.52)$$

from which it follows that the integral in (I.1.8.1) converges.

I.2 Some elementary properties of relative entropy

I.2.1 Let us consider that the random vectors ξ,η are taking values in the measurable space (X,S). Let us consider also another measurable space $(\bar{X},\bar{S})$ and let $f(x)$ be a S-measurable function defined on (X,S) with values in $(\bar{X},\bar{S})$. The compound functions

$$\xi(\omega) = f(\xi(\omega)),\ \bar{\eta}(\omega) = f(-(\omega)) \tag{I.2.1.1}$$

are random vectors with values in $(\bar{X},\bar{S})$.

Let

$$P_\xi(T) = P\{\omega;\xi(\omega)\ \varepsilon\ T\};\ P_\eta(T) = P\{\omega;\eta(\omega)\ \varepsilon\ T\};\ T\ \varepsilon\ S \tag{I.2.1.2}$$

$$P_{\bar{\xi}}(\bar{T}) = P\{\omega;\bar{\xi}(\omega)\ \varepsilon\ \bar{T}\};\ P_{\bar{\eta}}(\bar{T}) = P\{\omega;\bar{\eta}(\omega)\ \varepsilon\ T\};\ \bar{T}\ \varepsilon\ \bar{S} \tag{I.2.1.2'}$$

Theorem I.4 *For any random vectors ξ,η, and any function f,*

$$h(\bar{\xi}:\bar{\eta}) \leq h(\xi:\eta) \tag{I.2.1.3}$$

with equality if f^{-1} *exists a.e.*$(P_\xi + P_\eta)$.

Proof. Let $\bar{T}\ \varepsilon\ \bar{X}$, so that

$$T = f^{-1}(\bar{T}) = \{x\ \varepsilon\ X;\ f(x)\ \varepsilon\ \bar{T}\} \tag{I.2.1.4}$$

Because $f(x)$ is S-measurable, it follows that if $\bar{T}\ \varepsilon\ \bar{S}$, then $T = f^{-1}(\bar{T})\ \varepsilon\ S$ and

$$P_\xi(T) = P_\xi\{x;x\ \varepsilon\ T\} = P_{\bar{\xi}}\{\bar{x};\ \bar{x}\ \varepsilon\ \bar{T}\} = P_{\bar{\xi}}(\bar{T}) \tag{I.2.1.5}$$

$$P_\eta(T) = P_\eta\{x;x\ \varepsilon\ T\} = P_{\bar{\eta}}\{\bar{x};\ \bar{x}\ \varepsilon\ \bar{T}\} = P_{\bar{\eta}}(\bar{T}) \tag{I.2.1.5'}$$

From the relation

$$f^{-1}\left(\bigcup_{i=1}^{k} T_i\right) = \bigcup_{i=1}^{k} f^{-1}(\bar{T}_i) \tag{I.2.1.6}$$

if $\bar{T}_i\ \varepsilon\ \bar{S}$ $(1 \leq i \leq k)$ it follows that $T_i = f^{-1}(\bar{T}_i)$ $(1 \leq i \leq k)$; also if $\bigcup_{i=1}^{n} \bar{T}_i = \bar{X}$, it follows $\bigcup_{i=1}^{n} T_i = X$. Also from

$$f^{-1}(\bar{T}_1 - \bar{T}_2) = f^{-1}(\bar{T}_1) - f^{-1}(\bar{T}_2) \tag{I.2.1.7}$$

it follows that if $\bar{T}_1 \cap \bar{T}_2 = \emptyset$, then $T_1 \cap T_2 = \emptyset$.

From the above it follows that if $\bar{Z} = \{\bar{Z}_s\}$ is a partition of $\bar{X}$, then $Z = f^{-1}(\bar{Z}) = \{Z_s\}$ with $Z_s = f^{-1}(\bar{Z}_s)$ is a partition in X.

Let us denote the totality of such partitions $Z = f^{-1}(\bar{Z})$ if X by R, all partitions of X by U_S, all partitions of $\bar{X}$ by $U_{\bar{S}}$, so that $R \subset U_S$ and $U_S = R \cup (U_S - R)$.

From (I.2.1.5), (I.2.1.5') it follows

$$h(\bar{\xi}_{\bar{Z}} : \bar{\eta}_{\bar{Z}}) = \sum_{s=1}^{n} P_{\bar{\xi}}(\bar{Z}_s) \log \frac{P_{\bar{\xi}}(\bar{Z}_s)}{P_{\bar{\eta}}(\bar{Z}_s)} = \sum_{s=1}^{n} P_{\xi}(Z_s) \log \frac{P_{\xi}(Z_s)}{P_{\eta}(Z_s)} = h(\xi_Z : \eta_Z) \tag{I.2.1.8}$$

So

$$h(\bar{\xi} : \bar{\eta}) = \sup_{\bar{Z} \in U_{\bar{S}}} h(\bar{\xi}_{\bar{Z}} : \bar{\eta}_{\bar{Z}}) = \sup_{Z \in R} h(\xi_Z : \eta_Z) \tag{I.2.1.9}$$

and

$$h(\xi : \eta) = \sup_{Z \in U_S} h(\xi_Z : \eta_Z) = \max\{\sup_{Z \in R} h(\xi_Z : \eta_Z);\ \sup_{Z \in U_S - R} h(\xi_Z : \eta_Z)\}$$

$$= \max\{\sup_{\bar{Z} \in U_{\bar{S}}} h(\bar{\xi}_{\bar{Z}} : \bar{\eta}_{\bar{Z}});\ \sup_{Z \in U_S - R} h(\xi_Z : \eta_Z)\} = \max\{h(\bar{\xi} : \bar{\eta});\ \sup_{Z \in U_S - R} h(\xi_Z : \eta_Z)\} \geq h(\bar{\xi} : \bar{\eta}) \tag{I.2.1.10}$$

which proves (I.2.1.3). In the case when the inverse function f^{-1} exists $P_\xi + P_\eta$, a.e., we can change the roles of ξ, η with $\bar{\xi}, \bar{\eta}$, obtaining the inequality inverse to (I.2.1.3) so these both together give us the equality.

Obviously the above result remains true in the particular case when $(\bar{X}, \bar{S})$ is identical with (X,S); of particular interest is the case when f is a linear function, and let

$$\bar{\xi} = f(\xi) = A\xi, \quad \bar{\eta} = f(\eta) = A\eta.$$

<u>Theorem I.4'</u>. $h(A\xi : A\eta) \leq h(\xi : \eta)$ (I.2.1.11)

<u>with equality if A is not singular</u>.

This theorem is following from Theorem 4, but it can be obtained also from the fact that the integral representation (I.1.8.1) of $h(\xi : \eta)$ does not depend on the system of coordinates in the vector space X.

<u>Theorem I.5</u> <u>If ξ_n converges in distribution to ξ and η_n to η, then</u>

$$h(\xi:\eta) \leq \underline{\lim}_{n\to\infty} h(\xi_n:\eta_n) \qquad \text{(I.2.2.2)}$$

<u>Proof</u>. From Definition I.1 it follows that

1^o if $h(\xi:\eta) < \infty$, for any $\varepsilon > 0$ it exists a partition

$Z = \{Z_s\} \in U$, such that

$$h(\xi_Z:\eta_Z) > h(\xi:\eta) - \varepsilon \qquad \text{(I.2.2.3)}$$

2^o If $h(\xi:\eta) = \infty$, for any $N > 0$ it exists a partition

$Z = \{Z_s\} \in U$, such that

$$h(\xi_Z:\eta_Z) > N \qquad \text{(I.2.2.3')}$$

Becasue $Z \in U$, it follows that for any Z_s belonging to Z, in both cases 1^o, 2^o

$$\lim_{n\to\infty} P_{\xi_n}(Z_s) = P_\xi(Z_s); \ \lim_{n\to\infty} P_{\eta_n}(Z_s) = P_\eta(Z_s) (1 \leq s \leq m) \qquad \text{(I.2.2.4)}$$

From

$$h(\xi_{nZ}:\eta_{nZ}) = \sum_{i=1}^{m} P_{\xi_n}(Z_s) \log \frac{P_{\xi_n}(Z_s)}{P_{\eta_n}(Z_s)} \qquad \text{(I.2.2.5)}$$

$$h(\xi_Z:\eta_Z) = \sum_{i=1}^{m} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} \qquad \text{(I.2.2.5')}$$

it follows

$$\lim_{n\to\infty} h(\xi_{nZ}:\eta_{nZ}) = \sum_{i=1}^{m} [\lim_{n} P_{\xi_n}(Z_s)] \log \frac{[\lim_{n\to\infty} P_{\xi_n}(Z_s)]}{[\lim_{n\to\infty} P_{\eta_n}(Z_s)]} =$$

$$\sum_{i=1}^{m} P_\xi(Z_s) \log \frac{P_\xi(Z_s)}{P_\eta(Z_s)} = h(\xi_Z:\eta_Z) \qquad \text{(I.2.2.6)}$$

Obviously,

$$h(\xi_n:\eta_n) = \sup_{Z^{(0)}\varepsilon U_S} h(\xi_{nZ^{(0)}}:\eta_{nZ^{(0)}}) \geq h(\xi_{nZ}:\eta_{nZ}) \qquad \text{(I.2.2.7)}$$

From (I.2.2.6), (I.2.2.7) it follows

$$\underline{\lim}_{n\to\infty} h(\xi_n:\eta_n) \geq \lim_{n\to\infty} h(\xi_{nZ}:\eta_{nZ}) = h(\xi_Z:\eta_Z) \qquad \text{(I.2.2.8)}$$

i.e.

$$\underline{\lim}_{n\to\infty} h(\xi_n:\eta_n) \geq h(\xi_Z:\eta_Z) \qquad \text{(I.2.2.8')}$$

1° From (I.2.2.3), (I.2.2.8) it follows

$$\underline{\lim}_{n\to\infty} h(\xi_n:\eta_n) \geq h(\xi:\eta) - \varepsilon \qquad \text{(I.2.2.9)}$$

for any $\varepsilon > 0$.

2° From (I.2.2.3'), (I.2.2.8') it follows

$$\underline{\lim}_{n\to\infty} h(\xi_n:\eta_n) > N \qquad \text{(I.2.2.9')}$$

for any $N > 0$.

From (I.2.2.9) it follows (I.2.2.2) in the case $h(\xi:\eta) < \infty$ and from (I.2.2.9') it follows

$$\underline{\lim}_{n\to\infty} h(\xi_n:\eta_n) = \infty \qquad \text{(I.2.2.10)}$$

in the case $h(\xi:\eta) = \infty$, i.e. (I.2.2.2) So the theorem is proved.

I.2.3. Let us consider two sequences of random vectors ξ_n, η_n taking values in the measurable spaces (X_n, S_n) $(n = 1,2,\ldots)$ so that the random vectors $\xi^{(n)} = (\xi_1,\ldots,\xi_n)$, $\eta^{(n)} = (\eta_1,\ldots,\eta_n)$ are taking values in the measurable spaces

$$(X^{(n)}, S^{(n)}) = \mathop{X}_{i=1}^{n} (X_i, S_i) \qquad (n = 1,2,\ldots)$$

with elements $x^{(n)} \in X^{(n)}$, and the random vectors $\xi = (\xi_1,\ldots)$, $\eta = (\eta_1,\ldots)$ are taking values in the measurable space

$$(X,S) = \mathop{X}_{i=1} (X_i, S_i)$$

with elements $x \in X$. Obviously $x^{(n)} = (x_1,\ldots,x_n) \in X^{(n)}$, $x = (x_1,\ldots) \in X$

Theorem I.6

$$\lim_{n\to\infty} h(\xi^{(n)}:\eta^{(n)}) = h(\xi:\eta) \tag{I.2.3.1}$$

Proof. For any $x^{(n+1)} = (x_1,\ldots,x_n,x_{n+1}) \in X^{(n+1)}$ it corresponds an element $x^{(n)} = (x_1,\ldots,x_n) \in X^{(n)}$. Let us denote by F_n this correspondence, i.e.

$$F_n(x^{(n+1)}) = x^{(n)} \tag{I.2.3.2}$$

is a function with domain $X^{(n+1)}$ and range $X^{(n)}$. Consequently for $\xi^{(n)} = (\xi_1,\ldots,\xi_n)$, $\xi^{(n+1)} = (\xi_1,\ldots,\xi_n,\xi_{n+1})$ takes place the relation

$$F_n(\xi^{(n+1)}) = \xi^{(n)} \tag{I.2.3.3}$$

From (I.2.1.3), (I.2.3.3) it follows that

$$h(\xi^{(n+1)}:\eta^{(n+1)}) \geq h(\xi^{(n)}:\eta^{(n)}) \quad (n = 1,2,\ldots) \tag{I.2.3.4}$$

So that the sequence $h(\xi^{(n)}:\eta^{(n)})$ is not decreasing and consequently

$$\lim_{n\to\infty} h(\xi^{(n)}:\eta^{(n)}) \tag{I.2.3.5}$$

does exists.

Let us denote

1. M - the totality of all sets of the form

$$T = \mathop{X}_{i=1}^{n} Z_i \times \mathop{X}_{j=n+1}^{\infty} X_j \in S = \mathop{X}_{i=1}^{\infty} S_i \qquad (I.2.3.6)$$

where $Z_i \in S_i$, $Z_i \neq X_i$ $(1 \leq i \leq n)$. $(n = 1,2,\ldots)$

2. L - the algebra of all finite sums of sets belonging to M.

3. $r(T) = n$ for T in (I.2.3.6)

4. M_n - the totality of sets $T \in S$ of the form (I.2.3.6) with given $r(T) = n$.

5. L_n - the algebra of all finite sums of sets belonging to M_n.

6. U_L - the totality of partitions V of X with elements in L.

7. U_{L_n} - the totality of partitions V of X with elements in L_n.

8. U_M - the totality of partitions V of X with elements in M.

9. U_{M_n} - the totality of partitions V of X with elements in M_n.

10. $M^{(n)}$ - the totality of sets of the form

$$T^{(n)} = \mathop{X}_{i=1}^{n} Z_i \in S^{(n)} = \mathop{X}_{i=1}^{n} S_i \qquad (I.2.3.6')$$

11. $L^{(n)}$ - the algebra of all finite sums of sets belonging to $M^{(n)}$.

12. $U_{L^{(n)}}$ - the totality of partitions $V^{(n)}$ of $X^{(n)}$ with elements in $L^{(n)}$.

13. $U_{M^{(n)}}$ - the totality of partitions $V^{(n)}$ of $X^{(n)}$ with elements in $M^{(n)}$.

14. $$R = \bigcup_{n=1}^{\infty} U_{M_n}. \qquad (I.2.3.7)$$

It is obvious that

1°. L generates the σ-algebra S.

2°. Any partition $V \in U_L$ has a subpartition $V_0 \in U_M$.

3°. Any partition $V \in U_M$ has a subpartition $V_0 \in U_{M_n}$ for some value of n, because the number of elements in V is finite, i.e., any partition $V \in U_M$ has a subpartition $V_0 \in R$.

Consequently, from 2° and 3° it follows that any partition $V \in U_L$ has a subpartition $V_0 \in R$ and from Theorem I.2 it follows

$$h(\xi:\eta) = \sup_{V\in R} h(\xi_V \cdot \eta_V) = \sup_{1\le n<\infty} \sup_{V\in U_{M_n}} h(\xi_V:\eta_V) \tag{I.2.3.8}$$

4°. $L^{(n)}$ generates the σ-algebra $S^{(n)}$.

5°. Any partition $V^{(n)} \in U_{L^{(n)}}$ has a subpartition $V^{(n)}_{(0)} \in U_{M^{(n)}}$ so that from Theorem I.2

$$h(\xi^{(n)}:\eta^{(n)}) = \sup_{V^{(n)}\in U_{M^{(n)}}} h(\xi^{(n)}_{V^{(n)}}:\eta^{(n)}_{V^{(n)}}) \tag{I.2.3.9}$$

Let A_n be the one-to-one transformation established by (I.2.3.6), (I.2.3.6') between M and $M^{(n)}$, i.e.

$$A_n(T) = T^{(n)},\ T \in M,\ T^{(n)} \in M^{(n)} \tag{I.2.3.10}$$

Obviously, A_n is measure preserving in the sense that

$$P_{\xi^{(n)}}(T^{(n)}) = P_\xi(T),\ P_{\eta^{(n)}}(T^{(n)}) = P_\eta(T) \tag{I.2.3.11}$$

We can define a one-to-one correspondence B_n between U_{M_n} and $U_{M^{(u)}}$, so that

$$B_n(V) = V^{(n)},\ V \in U_{M_n},\ V^{(n)} \in U_{M^{(n)}} \tag{I.2.3.12}$$

where $V = \{V_s\}$, $V_s \in M_n$, $(1 \le s \le e)$, $V^{(n)} = \{V_s^{(n)}\}$, $V_s^{(n)} \in M^{(n)}$ $(1 \le s \le e)$, $V_s = V_s^{(n)} \times \mathop{X}\limits_{i=n+1}^{\infty} X_i$; i.e. $A_n(V_s) = V_s^{(n)} (1 \le s \le e)$.

Consequently for $V \in U_{M_n}$, $V^{(n)} = B_n(V) \in U_{M^{(n)}}$,

$$h(\xi_V:\eta_V) = \sum_{s=1}^{e} P_{\xi_V}(V_s) \log \frac{P_{\xi_V}(V_s)}{P_{\eta_V}(V_s)} =$$

$$= \sum_{s=1}^{e} P_{\xi^{(n)}_{V^{(n)}}}(V_s^{(n)}) \log \frac{P_{\xi^{(n)}_{V^{(n)}}}(V_s^{(n)})}{P_{\eta^{(n)}_{V^{(n)}}}(V_s^{(n)})} = h(\xi^{(n)}_{V^{(n)}}:\eta^{(n)}_{V^{(n)}}) \tag{I.2.3.13}$$

From (I.2.3.9), (I.2.3.11), (I.2.3.12), (I.2.3.13) it follows

$$\sup_{V\epsilon U_{M_n}} h(\xi_V:\eta_V) = \sup_{V^{(n)}\epsilon U_{M^{(n)}}} h(\xi^{(n)}_{V(n)}:\eta^{(n)}_{V(n)}) = h(\xi^{(n)}:\eta^{(n)}) \tag{I.2.3.14}$$

From (I.2.3.8) it follows

$$h(\xi:\eta) = \sup_{1\le n<\infty} h(\xi^{(n)}:\eta^{(n)}) \tag{I.2.3.15}$$

and from (I.2.3.4), (I.2.3.5) it follows

$$\sup_{1\le n<\infty} h(\xi^{(n)}:\eta^{(n)}) = \lim_{n\to\infty} h(\xi^{(n)}:\eta^{(n)}) \tag{I.2.3.16}$$

and from (I.2.3.15), (I.2.3.16) it follows (I.2.3.1).

I.2.4. If μ is a signed measure defined on all $T \epsilon S$, a set $E \epsilon S$ is <u>positive</u> with respect to μ if for any $T \epsilon S$

$$\mu(T \cap E) \geq 0 \tag{I.2.4.1}$$

and is <u>negative</u> with respect to μ if for any $T \epsilon S$

$$\mu(T \cap E) \leq 0 \tag{I.2.4.1'}$$

From this point of view, the empty set is both positive and negative.

It is known from Theorem A, §29, p. 121[4], that if μ is a signed measure, then there exists two disjoint sets X^+, $X^- \epsilon S$ such that their union is X and so that X^+ is positive and X^- is negative with respect to μ. They form a Hahn partition of X with respect to μ.

Let $Z = \{Z_s\}$ be a partition of X, so that the family of sets $Z^+ = \{Z_s^+\}$, where $Z_s^+ = Z_s \cap X^+$ is a partition of X^+, and the class $Z^- = \{Z_s^-\}$, where $Z_s^- = Z_s \cap X^-$ is a partition of X^-.

Let $S^+ = S \cap X^+$ be the σ-algebra of all S-measurable sets in X^+ and $S^- = S \cap X^-$ the σ-algebra of all S-measurable sets in X^-. Also let U_S be the totality of partitions of X with elements in S, U_{S^+} the totality of partitions of X^+ with elements in S^+, U_{S^-} the totality of partitions of X^- with elements in S^-.

Let

$$\mu^{+}(T) = \sup_{Z^{+} \varepsilon U_{S+}} \sum_{s} \mu(T \cap Z_s^{+}) = \mu(T \cap \bigcup_{s} Z_s^{+}) = \mu(T \cap X^{+}) \qquad (I.2.4.2)$$

$$\mu^{-}(T) = \sup_{Z^{-} \varepsilon U_{S-}} \sum_{s} \mu(T \cap Z_s^{-}) = -\mu(T \cap \bigcup_{s} Z_s^{-}) = -\mu(T \cap X^{-}) \qquad (I.2.4.3)$$

$$|\mu|(T) = \sup_{Z \varepsilon U_S} \sum_{s} |\mu(T \cap Z_s)| = \sup_{Z^{+} \varepsilon U_{S+}} \sum_{s} \mu(T \cap Z_s^{+}) -$$

$$- \sup_{Z^{-} \varepsilon U_{S-}} \sum_{s} (T \cap Z_s^{-}) = \mu^{+}(T) + \mu^{-}(T) \qquad (I.2.4.4.)$$

The set functions μ^{+}, μ^{-}, $|\mu|$ are named positive (or upper) variation of μ, negative (or lower) variation of μ, and total variation of μ. Each of them is a measure and

$$\mu(T) = \mu^{+}(T) - \mu^{-}(T) \qquad (I.2.4.5)$$

If μ is finite or σ-finite, so are his variations (See Theorem B, §29, p. 123[4]). In what follows, we will denote

$$||\mu|| = |\mu|(X) \qquad (I.2.4.5')$$

<u>Theorem I.7</u>. <u>If</u> μ_1, μ_2 <u>are measures on</u> (X,S), <u>it exists a set</u> $X_0 \varepsilon S$ <u>with</u> $\mu_2(X_0) = 0$, <u>and a non-negative S-measurable function a(x), such that for any</u> $T \varepsilon S$

$$|\mu_1 - \mu_2|(T) = \int_T |a(x) - 1| \mu_2(dx) + \mu_1(T \cap X_0) \qquad (I.2.4.6)$$

<u>If</u> μ_1 <u>is absolutely continuous with respect to</u> μ_2 <u>on X</u>, (which fact we will denote in the future by $\mu_1 << \mu_2$), <u>then</u> $X_0 = \emptyset$ <u>and</u>

$$a(x) = \frac{\mu_1(dx)}{\mu_2(dx)} \qquad (I.2.4.7)$$

<u>so that</u>

$$|\mu_1 - \mu_2|(T) = \int_T |a(x) - 1| \mu(dx) \qquad (I.2.4.8)$$

If $\mu_1 \ll \nu$, $\mu_2 \ll \nu$ *and* $\pi_1(x)$, $\pi_2(x)$ *are the corresponding densities, then*

$$a(x) = \frac{\pi_1(x)}{\pi_2(x)} \qquad \text{(I.2.4.7')}$$

and

$$|\mu_1 - \mu_2|(T) = \int_T |\pi_1(x) - \pi_2(x)| \, \nu(dx) \qquad \text{(I.2.4.8')}$$

Proof. It is known that for any arbitrary measures μ_1, μ_2, $\mu = \mu_1 - \mu_2$ is a signed measure. It is also known that it exists a set $X_0 \in S$ such that $\mu_1 \ll \mu_2$ on $X - X_0$ and

$$\mu_1(T_0) = \int_{T_0} a(x) \, \mu_2(dx) \qquad \text{(I.2.4.9)}$$

where $T_0 \subset X - X_0$ and $a(x)$ is an S-measurable function.

Consequently, for $T_0 \subset X - X_0$

$$(\mu_1 - \mu_2)(T_0) = \mu_1(T_0) - \mu_2(T_0) = \int_{T_0} [a(x) - 1]\mu_2(dx) \qquad \text{(I.2.4.10)}$$

Because the total variation of $\mu_1 - \mu_2$ is a measure on S, it follows that for $T \in S$

$$|\mu_1 - \mu_2|(T) = |\mu_1 - \mu_2|[T \cap (X - X_0)] + |\mu_1 - \mu_2|(T \cap X_0) \qquad \text{(I.2.4.11)}$$

Now we will calculate the two elements in this sum.

The first element is

$$|\mu_1 - \mu_2|[T \cap (X - X_0)] = (\mu_1 - \mu_2)^+ [T \cap (X - X_0)] + (\mu_1 - \mu_2)^- [T \cap (X - X_0)] =$$

$$= (\mu_1 - \mu_2)[T \cap (X - X_0) \cap X^+] - (\mu_1 - \mu_2)[T \cap (X - X_0) \cap X^-] =$$

$$= \int_{T \cap (X - X_0) \cap X^+} [a(x) - 1]\mu_2(dx) - \int_{T \cap (X - X_0) \cap X^-} [a(x) - 1]\mu_2(dx) =$$

$$= \int_{T \cap (X - X_0) \cap X^+} |a(x) - 1|\mu_2(dx) + \int_{T \cap (X - X_0) \cap X^-} |a(x) - 1|\mu_2(dx) =$$

$$= \int_{T \cap (X - X_0)} |a(x) - 1|\mu_2(dx) = \int_{T \cap (X - X_0)} |a(x) - 1|\mu_2(dx) + \int_{T \cap X_0} |a(x) - 1|\mu_2(dx) =$$

$$= \int_T |a(x) - 1|\mu_2(dx) \qquad \text{(I.2.4.12)}$$

The second element is

$$|\mu_1 - \mu_2|(T \cap X_0) = (\mu_1 - \mu_2)^+(T \cap X_0) - (\mu_1 - \mu_2)^-(T \cap X_0) =$$

$$= (\mu_1 - \mu_2)[(T \cap X_0) \cap X^+] + (\mu_1 - \mu_2)[(T \cap X_0) \cap X^-] =$$

$$= \mu_1[(T \cap X_0) \cap X^+] + \mu_1[(T \cap X_0) \cap X^-] = \mu_1[(T \cap X_0) \cup (X^+ \cup X^-)] =$$

$$= \mu_1[(T \cap X_0) \cap X] = \mu_1(T \cap X_0) = \tag{I.2.4.13}$$

From (I.2.4.7'), (I.2.4.8) it follows (I.2.4.8'). So our theorem is proved.

Let P_ξ, P_η be given. If $P_\xi << P_\eta$, from Theorem I.3 formula (I.1.8.1) it follows

$$h(\xi:\eta) = \int_X a_{\xi:\eta}(x) \log a_{\xi:\eta}(x).P_\eta(dx) \tag{I.2.4.14}$$

and from Theorem I.7, formula (I.2.4.8) and (I.2.4.5') it follows

$$||P_\xi - P_\eta|| = \int_X |a_{\xi:\eta}(x) - 1|.P_\eta(dx) \tag{I.2.4.15}$$

where

$$a_{\xi:\eta}(x) = \frac{P_\xi(dx)}{P_\eta(dx)} \tag{I.2.4.16}$$

Theorem I.8.

a) For arbitrary random vectors ξ, η, takes place the inequality

$$||P_\xi - P_\eta||^2 \leq 2.h(\xi:\eta) \tag{I.2.4.17}$$

b) For arbitrary small $\delta > 0$, there exists random vectors ξ, η such that

$$||P_\xi - P_\eta||^2 > (2 - \delta)\ h(\xi:\eta) \tag{I.2.4.17'}$$

so in (I.2.4.17) the constant 2 can not be replaced by a smaller one.

Proof. In the case when P_ξ is not absolute continuous with respect to P_η, the second member in (I.2.4.17) is not finite, so (I.2.4.17) is trivially true, so it remains to be proved only in the case when $P_\xi << P_\eta$. In this case

$$P_\xi(Z) = \int_Z a_{\xi:\eta}(x).P_\eta(dx) \tag{I.2.4.18}$$

a) Let us consider the function

$$\psi(z) = z \log z - z + 1 \qquad (0 \leq z < \infty) \tag{I.2.4.19}$$

Because $\psi'(z) = \log z$, $\psi''(z) = \frac{1}{z}$, it is easy to see that for $z = 1$ this function has a minimum $\psi(1) = 0$, and is convex, non-negative, so that

$$\psi(z) \geq 0 \qquad (0 \leq z < \infty) \tag{I.2.4.19'}$$

Let us consider the expression

$$\phi(z) = \frac{2}{3}(2 + z)\,\psi(z) - (z-1)^2 = \frac{2}{3}(2+z)(z \log z - z + 1)^2 - (z-1)^2 \tag{I.2.4.20}$$

Because

$$\phi(1) = 0 \tag{I.2.4.21}$$

$$\phi'(z) = \frac{2}{3}(z \log z - z + 1) + \frac{2}{3}(2+z) \log z - 2(z-1) \tag{I.2.4.22}$$

$$\phi'(1) = 0 \tag{I.2.4.23}$$

$$\phi''(z) = \frac{4}{3z}\,(z \log z - z + 1) = \frac{4}{3z}.\psi(z) \tag{I.2.4.24}$$

it follows that the function $\phi(z)$ has a minimum for $z = 1$ and $\phi(1) = 0$, and from (I.2.4.19') it follows from (I.2.4.24) that

$$\phi''(z) \geq 0 \; (z \geq 0) \tag{I.2.4.25}$$

i.e. $\phi(z)$ is convex, and consequently

$$\phi(z) \geq 0 \quad (z \geq 0) \tag{I.2.4.25'}$$

i.e.

$$(z-1)^2 \leq \frac{2}{3}\,(2+z)(z \log z - z + 1) \quad (z \geq 0) \tag{I.2.4.26}$$

or

$$(z-1)^2 \leq \frac{2}{3}\,(2+z)\,\psi(z) \qquad (z \geq 0) \tag{I.2.4.27}$$

or

$$|z-1| \leq (\frac{4}{3} + \frac{2}{3}z)\,\psi^{\frac{1}{2}}(z) \quad (z \geq 0) \tag{I.2.4.28}$$

Replacing in (I.2.4.28) z with $a_{\xi:\eta}(x)$, we obtain

$$|a_{\xi:\eta}(x) - 1| \leq (\frac{4}{3} + \frac{2}{3}\,a_{\xi:\eta}(x))^{\frac{1}{2}}.\psi^{\frac{1}{2}}\,(a_{\xi:\eta}(x)) \tag{I.2.4.29}$$

from which it follows

$$\int_X |a_{\xi:\eta}(x) - 1| . P_\eta(dx) \leq \int_X (\frac{4}{3} + \frac{2}{3}\, a_{\xi:\eta}(x))^{\frac{1}{2}} . \psi^{\frac{1}{2}}(a_{\xi:\eta}(x)) . P_\eta(dx) \qquad (I.2.4.30)$$

From Cauchy-Schwartz's inequality, we obtain

$$||P_\xi - P_\eta||^2 = (\int_X |a_{\xi:\eta}(x) - 1| . P_\eta(dx))^2 \leq$$

$$\leq (\int_X (\frac{4}{3} + \frac{2}{3}\, a_{\xi:\eta}(x))^{\frac{1}{2}} . \psi^{\frac{1}{2}}(a_{\xi:\eta}(x)) P_\eta(dx))^2 \leq$$

$$\leq \int_X (\frac{4}{3} + \frac{2}{3}\, a_{\xi:\eta}(x)) . P_\eta(dx) . \int_X \psi(a_{\xi:\eta}(x)) . P_\eta(dx) =$$

$$= [\frac{4}{3} . \int_X P_\eta(dx) + \frac{2}{3} \int_X a_{\xi:\eta}(x) . P_\eta(dx)] \cdot$$

$$\cdot [\int_X a_{\xi:\eta}(x) \log a_{\xi:\eta}(x) . P_\eta(dx) - \int_X a_{\xi:\eta}(x) . P_\eta(dx) + \int_X P_\eta(dx)] =$$

$$= (\frac{4}{3} + \frac{2}{3}) [\int_X a_{\xi:\eta}(x) \log a_{\xi:\eta}(x) . P_\eta(dx) - 1 + 1] =$$

$$= 2 \int_X a_{\xi:\eta}(x) \log a_{\xi:\eta}(x) . P_\eta(dx) = 2\, h(\xi:\eta) \qquad (I.2.4.31)$$

which proves (I.2.4.17).

b) Let ξ be a random variable, such that it exists a set $Z_0 \varepsilon S$ with

$$P_\xi(Z_0) = P_\xi(X - Z_0) = \frac{1}{2} \qquad (I.2.4.32)$$

Because for any $Z \varepsilon S$

$$P_\xi(Z) = \int_Z 1 . P_\xi(dx) \qquad (I.2.4.33)$$

it follows

$$\pi_\xi(x) = 1 \qquad (x \varepsilon X) \qquad (I.2.4.34)$$

Let δ be an arbitrary number $(0 < \delta < 1)$, and let us define the function

$$\pi_\eta(x) = 1 - \delta \qquad x \varepsilon Z_0 \qquad (I.2.4.35)$$

$$\pi_\eta(x) = 1 + \delta \qquad x \varepsilon X Z_0 \qquad (I.2.4.36)$$

Because $\pi_\eta(x) > 0$ ($x \in X$) and

$$\int_X \pi_\eta(x)\ P_\xi(dx) = 1 \qquad (I.2.4.37)$$

it follows that $\pi_\eta(x)$ is the probability density of some random variable with respect to the measure P_ξ.

Consequently

$$h(\xi:\eta) = \int_X \pi_\eta(x)\ \log \frac{\pi_\xi(x)}{\pi_\eta(x)}\ .P_\xi(dx) = \int_{Z_0} \log \frac{1}{1-\delta}\ P_\xi(dx) + \int_{X-Z_0} \log \frac{1}{1+\delta}\ P_\xi(dx)$$

$$= \frac{1}{2} \log \frac{1}{1-\delta} + \frac{1}{2} \log \frac{1}{1+\delta} = \frac{1}{2} \log \frac{1}{1-\delta^2} \qquad (I.2.3.38)$$

Let us consider the function

$$g(z) = \log\ (1 - z)^{-1} \qquad (I.2.4.39)$$

It is easy to calculate that

$$\frac{d^n g}{dz^n} = (n - 1)\,!\ (1 - z)^{-n} \quad (n \geq 1) \qquad (I.2.4.40)$$

So that

$$g(0) = 0,\ g^{(n)}(0) = (n - 1)\,! \quad (n \geq 1) \qquad (I.2.4.41)$$

and consequently,

$$g(z) = \sum_{k=1}^{\infty} \frac{z^k}{k} \qquad (I.2.4.42)$$

a convergent series for $|z| < 1$; because $\delta^2 < \delta < 1$, and from (I.2.4.38) we know that

$$h(\xi:\eta) = \frac{1}{2}\, g(\delta^2) \qquad (I.2.4.43)$$

and from (I.2.4.42) it follows

$$h(\xi:\eta) = \frac{1}{2} \sum_{k=1}^{\infty} \frac{\delta^{2k}}{k} \qquad (I.2.4.44)$$

i.e.

$$2\ h(\xi:\eta) = \delta^2 + \sum_{k=2}^{\infty} \frac{\delta^{2k}}{k} \qquad (I.2.4.45)$$

Also from (I.2.4.5'), (I.2.4.8') it follows that

$$||P_\xi - P_\eta|| = \int_X |\pi_\xi(x) - \pi_\eta(x)|\,.P_\xi(dx) = \int_{Z_0} |1 - (1 - \delta)|\,.P_\xi(dx) + \int_{X-Z_0} |1 - (1 + \delta)|\,.P_\xi(dx) =$$

$$= \frac{\delta}{2} + \frac{\delta}{2} = \delta \qquad (I.2.4.46)$$

From (I.2.4.45), (I.2.4.47) it follows that

$$2\, h(\xi:\eta) = ||P_\xi - P_\eta||^2 + \sum_{k=2}^{\infty} \frac{\delta^{2k}}{k} \tag{I.2.4.47}$$

So, from (I.2.4.46) we obtain

$$(2-\delta)\, h(\xi:\eta) - ||P_\xi - P_\eta||^2 = (2-\delta)\frac{1}{2}(\delta^2 + \sum_{k=2}^{\infty} \frac{\delta^{2k}}{k}) - \delta^2 =$$

$$= (1 - \frac{\delta}{2}) \cdot \sum_{k=2}^{\infty} \frac{\delta^{2k}}{k} - \frac{\delta^3}{2} \tag{I.2.4.48}$$

If δ is sufficient small

$$(1 - \frac{\delta}{2}) \sum_{k=2}^{\infty} \frac{\delta^{2k}}{k} < \frac{\delta^3}{4} \tag{I.2.4.49}$$

so that from (I.2.4.48) it follows

$$(2-\delta)\, h(\xi:\eta) - ||P_\xi - P_\eta||^2 < \frac{\delta^3}{4} - \frac{\delta^3}{2} = -\frac{\delta^3}{4} < 0 \tag{I.2.4.50}$$

Consequently for ξ,η as defined above, (I.2.4.17) is satisfied.

I.2.5. Theorem I.9.

a) If ξ,η are given random vectors and $\rho = \rho_0$ $(0 < \rho_0 < 1)$ satisfies the relation

$$h(\xi:\eta) + h(\xi:\eta) = 2\rho \log \frac{1-\rho}{1+\rho} \tag{I.2.5.1}$$

then

$$||P_\xi - P_\eta|| \leq 2\rho_0 \tag{I.2.5.2}$$

b) Between all pairs of vectors with the same value of

$$h(\xi:\eta) + h(\xi:\eta) \tag{I.2.5.1'}$$

it exists a pair ξ_0,η_0 such that

$$||P_{\xi_0} - P_{\eta_0}|| = 2\rho_0 \tag{I.2.5.2'}$$

so that the relation (I.2.5.2) cannot be improved.

Proof.

a_1) It is easy to see that

$$F(\rho) = 2\rho \log \frac{1+\rho}{1-\rho} \quad (0 \leq \rho < 1) \tag{I.2.5.3}$$

has the derivative

$$F'(\rho) = \frac{4\rho}{1-\rho^2} + 2\log\frac{1+\rho}{1-\rho} \qquad (I.2.5.3')$$

Because $F(0) = 0$, $F'(\rho) \geq 0$ $(0 \leq \rho < 1)$, it follows that $F(\rho)$ in the interval $0 \leq \rho < 1$ is monotonously increasing and has range $[0, +\infty)$. More than that, it can be written as

$$F(\rho) = 4.\sum_{m=1}^{\infty} \frac{\rho^{2m}}{2m-1} \qquad (I.2.5.4)$$

from where it is also seen that it is increasing. Consequently, for any value of the constant J, the equation

$$2\rho \log\frac{1+\rho}{1-\rho} = J \quad (0 \leq J < \infty) \qquad (I.2.5.5)$$

has exactly one solution.

$$\rho = \rho(J), \qquad 0 \leq \rho(J) < 1 \qquad (I.2.5.5')$$

Given the random vectors ξ, η, we define the value of J_0 by

$$J_0 = h(\xi:\eta) + h(\xi:\eta) \qquad (I.2.5.6)$$

a_2) Denoting by

$$\rho_0 = \rho(J_0) \qquad (I.2.5.6')$$

the solution of the equation (I.2.5.5) with $J = J_0$, let us denote

$$\sigma = \frac{2\rho_0^2}{1-\rho_0^2}, \quad \tau = \frac{J_0+2\sigma}{2\rho_0} = \log\frac{1+\rho_0}{1-\rho_0} + \frac{2\rho_0}{1-\rho_0^2} \qquad (I.2.5.7)$$

We will prove that the inequality

$$\tau\,|z-1| \leq (z-1)\log z + \sigma\,(z+1) \qquad (I.2.5.8)$$

is satisfied for all $z \geq 0$.

If (I.2.5.8) is satisfied for a value of z, it is satisfied also for $\frac{1}{z}$, because

$$\tau\,\left|\frac{1}{z} - 1\right| \leq \left(\frac{1}{z} - 1\right)\log\frac{1}{z} + \sigma\left(\frac{1}{z} + 1\right) \qquad (I.2.5.8')$$

is another manner of writing (I.2.5.8).

If $z < 1$ i.e. $z - 1 < 0$, then obviously,

$$\frac{1}{z} - 1 = \frac{z-1}{z} > 0.$$

Consequently, in order to verify (I.2.5.8) for $z \geq 0$, it is sufficient to verify

$$\tau(z-1) \leq (z-1)\log z + \sigma(z+1) \qquad \text{(I.2.5.8'')}$$

for $z \leq 1$.

Let

$$\phi(z) = (z-1)\log z + (\sigma-\tau)z + \sigma + \tau \qquad \text{(I.2.5.9)}$$

From (I.2.5.7) it is easy to calculate

$$\sigma - \tau = -\frac{2\rho_0}{1+\rho_0} - \log\frac{1+\rho_0}{1-\rho_0};\ \sigma + \tau = \frac{2\rho_0}{1-\rho_0} + \log\frac{1+\rho_0}{1-\rho_0} \qquad \text{(I.2.5.10)}$$

so that (I.2.5.8'') takes the form

$$\phi(z) = (z-1)\log z + \left(-\frac{2\rho_0}{1+\rho_0} - \log\frac{1+\rho_0}{1-\rho_0}\right)z + \left(\frac{2\rho_0}{1-\rho_0} + \log\frac{1+\rho_0}{1-\rho_0}\right) \qquad \text{(I.2.5.11)}$$

so that

$$\phi'(z) = \log z - \frac{1}{z} + \frac{1-\rho_0}{1+\rho_0} - \log\frac{1+\rho_0}{1-\rho_0} \qquad \text{(I.2.5.11')}$$

$$\phi''(z) = \frac{1}{z} + \frac{1}{z^2} > 0 \quad (z \geq 0) \qquad \text{(I.2.5.11'')}$$

Let us find a solution z_0 of the system of the two equations

$$\phi(z) = 0, \quad z\phi'(z) = 0 \qquad \text{(I.2.5.12)}$$

The equations can be written as

$$(z-1)\log\left[z\left(\frac{1+\rho_0}{1-\rho_0}\right)^{-1}\right] - \frac{2\rho_0}{1+\rho_0}\left(z - \frac{1+\rho_0}{1-\rho_0}\right) = 0 \qquad \text{(I.2.5.13)}$$

$$z\log\left[z\left(\frac{1+\rho_0}{1-\rho_0}\right)^{-1}\right] + \frac{1+\rho_0}{1-\rho_0}\left(z - \frac{1+\rho_0}{1-\rho_0}\right) = 0 \qquad \text{(I.2.5.13')}$$

and it is obvious that they admit the common root

$$z = \frac{1+\rho_0}{1-\rho_0} > 0 \qquad \text{(I.2.5.14)}$$

Because the function $\phi(z)$ has a minimum zero for z given by (I.2.5.14), because

$$\phi(+0) = \infty, \quad \phi(+\infty) = +\infty \tag{I.2.5.15}$$

and because from (I.2.5.11'') it is seen that $\phi(z)$ is convex for $z \geq 0$, it follows that $\phi(z) \geq 0$ $(z \geq 0)$, i.e. the inequality (I.2.5.8'') is satisfied, i.e. the inequality (I.2.5.8) is satisfied.

a_3) Let us replace in (I.2.5.8) z with $a_{\xi:\eta}(x)$, so we obtain

$$\tau|a_{\xi:\eta}(x) - 1| \leq (a_{\xi:\eta}(x) - 1)\log a_{\xi:\eta}(x) + \sigma(a_{\xi:\eta}(x) + 1) \tag{I.2.5.16}$$

and by integration on P_η over all X,

$$\tau . \int_X |a_{\xi:\eta}(x) - 1| . P_\eta(dx) \leq \int_X (a_{\xi:\eta}(x) - 1)\log a_{\xi:\eta}(x) . P_\eta(dx) +$$

$$+ \sigma \int_X (a_{\xi:\eta}(x) + 1) . P_\eta(dx) \tag{I.2.5.17}$$

Now it is easy to calculate

$$\int_X (a_{\xi:\eta}(x) - 1)\log a_{\xi:\eta}(x)\, P_\eta(dx) = \int_X a_{\xi:\eta}(x)\log A_{\xi:\eta}(x) P_\eta(dx) -$$

$$- \int_X \log a_{\xi:\eta}(x)\, P_\eta(dx) = h(\xi:\eta) + \int_X \frac{P_\eta(dx)}{P_\xi(dx)} \log a_{\eta:\xi}(x) . P_\xi(dx)$$

$$= h(\xi:\eta) + \int_X a_{\eta:\xi}(x)\log a_{\eta:\xi}(x) . P_\xi(dx) = h(\xi:\eta) + h(\eta:\xi) = J_0 \tag{I.2.5.18}$$

Also

$$\int_X (a_{\xi:\eta}(x) + 1)\, P_\eta(dx) = \int_X a_{\xi:\eta}(x)\, P_\eta(dx) + \int_X P_\eta(dx) =$$

$$= \int_X \frac{P_\xi(dx)}{P_\eta(dx)} . P_\eta(dx) + 1 = \int_X P_\xi(dx) + 1 = 2 \tag{I.2.5.19}$$

From (I.2.4.15) we have

$$\int_X |a_{\xi:\eta}(x) - 1|\, P_\eta(dx) = ||P_\xi - P_\eta|| \tag{I.2.5.20}$$

With the results (I.2.5.18), (I.2.5.19), (I.2.5.20), the inequality (I.2.5.17) can be written as

$$\tau \; ||P_\xi - P_\eta|| \leq J_0 + 2\sigma \qquad \text{(I.2.5.21)}$$

and because of (I.2.5.7)

$$||P_\xi - P_\eta|| \leq \frac{J_0 + 2\sigma}{\tau} = 2\rho_0 \qquad \text{(I.2.5.21')}$$

i.e. (I.2.5.2).

b) Let us consider two random variables ξ_0, η_0, both taking the values x_1, x_2 only, with probability given by

$$P_{\xi_0}(x_1) = P_{\eta_0}(x_2) = \frac{1}{2}\,(1+\rho_0);\; P_{\xi_0}(x_2) = P_{\eta_0}(x_1) = \frac{1}{2}\,(1-\rho_0) \qquad \text{(I.2.5.22)}$$

Then

$$h(\xi_0 : \eta_0) = P_{\xi_0}(x_1) \log \frac{P_{\xi_0}(x_1)}{P_{\eta_0}(x_1)} + P_{\xi_0}(x_2) \log \frac{P_{\xi_0}(x_2)}{P_{\eta_0}(x_2)}$$

$$= \rho_0 \log \frac{1+\rho_0}{1-\rho_0} \qquad \text{(I.2.5.23)}$$

so that

$$J_0 = 2\rho_0 \log \frac{1+\rho_0}{1-\rho_0} \qquad \text{(I.2.5.24)}$$

which proves the last statement of the theorem.

I.2.6. Theorem I.10.

If ξ, η are two random vectors and $\varepsilon > 0$ arbitrary, then

$$P_\xi\{|i_{\xi:\eta}(x)| > \varepsilon\} \leq ||P_\xi - P_\eta|| \, \frac{1+\varepsilon}{\varepsilon} \qquad \text{(I.2.6.1)}$$

with $i_{\xi:\eta}(x)$ given by (I.1.8.4).

Proof. Let

$$A = \{x;\; |\log a_{\xi:\eta}(x)| = |i_{\xi:\eta}(x)| > \varepsilon\} \qquad \text{(I.2.6.2)}$$

$$A_1 = \{x;\; \log a_{\xi:\eta}(x) > \varepsilon\};\; A_2 = \{x;\; \log a_{\xi:\eta}(x) < -\varepsilon\} \qquad \text{(I.2.6.3)}$$

So

$$A = A_1 \cup A_2, \quad A_1 \cap A_2 = \emptyset \qquad \text{(I.2.6.4)}$$

Because

$$\varepsilon \geq \log(1+\varepsilon) \tag{I.2.6.5}$$

it follows that

$$A_1 = \{x;\ \log a_{\xi:\eta}(x) > \varepsilon\} \subset \{x;\ \log a_{\xi:\eta}(x) > \log(1+\varepsilon)\} =$$

$$= \{x;\ a_{\xi:\eta}(x) > 1+\varepsilon\} = \{x;\ \frac{1}{a_{\xi:\eta}(x)} < \frac{1}{1+\varepsilon}\} =$$

$$= \{x;\ 1 - \frac{1}{a_{\xi:\eta}(x)} > 1 - \frac{1}{1+\varepsilon}\} = \{x;\ 1 - \frac{1}{a_{\xi:\eta}(x)} > \frac{\varepsilon}{1+\varepsilon}\} \tag{I.2.6.6}$$

Because (I.2.6.5) can be written as

$$e^{\varepsilon} > 1 + \varepsilon \tag{I.2.6.7}$$

and

$$-\varepsilon < -\frac{\varepsilon}{1+\varepsilon} \tag{I.2.6.7'}$$

it follows that

$$A_2 = \{x;\ \log a_{\xi:\eta}(x) < -\varepsilon\} = \{x;\ \varepsilon < \log \frac{1}{a_{\xi:\eta}(x)}\} =$$

$$= \{x;\ \frac{1}{a_{\xi:\eta}(x)} > e^{\varepsilon}\} \subset \{x;\ \frac{1}{a_{\xi:\eta}(x)} > 1+\varepsilon\} =$$

$$= \{x;\ -\varepsilon > 1 - \frac{1}{a_{\xi:\eta}(x)}\} \subset \{x;\ 1 - \frac{1}{a_{\xi:\eta}(x)} < -\frac{\varepsilon}{1+\varepsilon}\} \tag{I.2.6.8}$$

From (I.2.6.4), (I.2.6.6), (I.2.6.8) it follows that

$$A = A_1 \cup A_2 \subset \{x;\ 1 - \frac{1}{a_{\xi:\eta}(x)} > \frac{\varepsilon}{1+\varepsilon}\} \cup \{x;\ 1 - \frac{1}{a_{\xi:\eta}(x)} < -\frac{\varepsilon}{1+\varepsilon}\} =$$

$$= \{x;\ \left|1 - \frac{1}{a_{\xi:\eta}(x)}\right| > \frac{\varepsilon}{1+\varepsilon}\} = K \tag{I.2.6.9}$$

From (I.2.6.2), (I.2.6.9) it follows

$$P_\xi\{x;\ |i_{\xi:\eta}(x)| > \varepsilon\} \leq P_\xi(K) \tag{I.2.6.10}$$

Let

$$N = \{x;\ a_{\xi:\eta}(x) = 0\}, \quad X - N = \{x;\ a_{\xi:\eta}(x) \neq 0\} \tag{I.2.6.11}$$

so that

$$P_\xi(N) = \int_N a_{\xi:\eta}(x)\, P_\xi(dx) = 0 \tag{I.2.6.12}$$

Let us denote

$$K_1 = K \cap (X-N) \tag{I.2.6.13}$$

so that

$$K = (K \cap N) \cup K_1 \tag{I.2.6.14}$$

From (I.2.6.12), (I.2.6.13), (I.2.6.14) it follows that

$$P_\xi(K \cap N) = 0 \tag{I.2.6.15}$$

$$P_\xi(K) = P_\xi(K \cap N) + P_\xi(K_1) = P_\xi(K_1) \tag{I.2.6.15'}$$

because from (I.2.6.12) and $K \subset N$ it follows

$$P_\xi(K) \leq P(N) = 0 \tag{I.2.6.12'}$$

Taking into consideration (I.2.6.10), (I.2.6.15') it follows

$$||P_\xi - P_\eta|| = \int_X |a_{\xi:\eta}(x) - 1| . P_\eta(dx) > \int_{K_1} |a_{\xi:\eta}(x) - 1| . P_\eta(dx) =$$

$$= \int_{K_1} |a_{\xi:\eta}(x) - 1| . \frac{P_\xi(dx)}{a_{\xi:\eta}(x)} = \int_{K_1} \left|1 - \frac{1}{a_{\xi:\eta}(x)}\right| . P_\xi(dx) \geq \int_{K_1} \frac{\varepsilon}{1+\varepsilon} . P_\xi(dx) =$$

$$= \frac{\varepsilon}{1+\varepsilon} . P_\xi(K_1) = \frac{\varepsilon}{1+\varepsilon} . P_\xi(K) \geq \frac{\varepsilon}{1+\varepsilon} . P\{x;\ |i_{\xi:\eta}(x)| > \varepsilon\} \tag{I.2.6.16}$$

i.e. (I.2.6.1).

I.3. Additivity theorems

I.3.1. Let (Ω_i,Σ_i,P_i) be a probability space, where Ω_i is a set of elements ω_i; Σ_i a σ-algebra of subsets of Ω_i; P_i - a probability measure on Σ_i $(i = 1,2)$

We consider the random vectors ξ_i, η_i $(i = 1,2)$, defined on these probability spaces, with values in the measurable spaces (X_i,S_i), where X_i is a set of elements x_i, S_i a σ-algebra of subsets of X_i.

Let $(\Omega,\Sigma) = (\Omega_1,\Sigma_1) \times (\Omega_2,\Sigma_2)$ be a measurable space, where Ω is the set of elements $\omega = (\omega_1,\omega_2)$, $\omega_1 \in \Omega_1$, $\omega_2 \in \Omega_2$, and $\Sigma = \Sigma_1 \times \Sigma_2$.

Let P be a probability measure on Σ with marginals P_i on Σ_i $(i = 1,2)$.

So (Ω,Σ,P) is a probability space.

We consider the random vectors

$$(\xi_1(\omega_1),\ \xi_2(\omega_2)), \quad (\eta_1(\omega_1),\ \eta_2(\omega_2)) \qquad \text{(I.3.1.1)}$$

defined on the probability space (Ω,Σ,P) with values in the measurable space $(X,S) = (X_1,S_1) \times (X_2,S_2)$, where X is the set of elements $x = (x_1,x_2)$, $x_1 \in X_1$, $x_2 \in X_2$ and $S = S_1 \times S_2$.

Let

$$P_{\xi_i}(T_i) = P_i\{\omega_i;\ \xi_i(\omega_i) \in T_i\},\ T_i \in S_i \quad (i = 1,2) \qquad \text{(I.3.1.2)}$$

$$P_{\eta_i}(T_i) = P_i\{\omega_i;\ \eta_i(\omega_i) \in T_i\},\ T_i \in S_i \quad (i = 1,2) \qquad \text{(I.3.1.3)}$$

be the probability measures of the random vectors ξ_i, η_i $(i = 1,2)$ and let

$$P_{\xi_1,\xi_2}(T) = P\{(\omega_1,\omega_2);\ (\xi_1(\omega_1),\ \xi_2(\omega_2)) \in T\},\ T \in S \qquad \text{(I.3.1.4)}$$

$$P_{\eta_1,\eta_2}(T) = P\{(\omega_1,\omega_2);\ (\eta_1(\omega_1),\ \eta_2(\omega_2)) \in T\},\ T \in S \qquad \text{(I.3.1.5)}$$

be their joint probability measures.

Lemma I.5.

If $P_{\xi_1 \xi_2}$ is absolutely continuous with respect to $P_{\eta_1 \eta_2}$, then

a) P_{ξ_1} is absolutely continuous with respect to P_{η_1}, P_{ξ_2} is absolutely continuous with respect to P_{η_2}, and

b) $P_{\xi_2|\xi_1}(.|x_1)$ is absolutely continuous with respect to $P_{\eta_2|\eta_1}(.|x_1)$ (P_{η_1} - a.e).

c) Moreover, if the Radon-Nicodym derivative of P_{ξ_1,ξ_2} with respect to P_{η_1,η_2} is

$$a_{(\xi_1\xi_2)\,:\,(\eta_1\eta_2)}(x_1,x_2) = \frac{P_{\xi_1\xi_2}(dx_1dx_2)}{P_{\eta_1\eta_2}(dx_1dx_2)} \qquad \text{(I.3.1 6)}$$

then the Radon-Nicodym derivative of P_{ξ_1} with respect to P_{η_1} is

$$a_{\xi_1:\eta_1}(x_1) = \frac{P_{\xi_1}(dx_1)}{P_{\eta_1}(dx_1)} = \int_{X_2} a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\; P_{\eta_2|\eta_1}(dx_2|x_1) \qquad \text{(I.3.I.7)}$$

P_{η_1} - a.e.), the Radon-Nicodym derivative of P_{ξ_2} with respect to P_{η_2} is

$$a_{\xi_2:\eta_2}(x_2) = \frac{P_{\xi_2}(dx_2)}{P_{\eta_2}(dx_2)} = \int_{X_1} a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\; P_{\eta_1|\eta_2}(dx_1|x_2) \qquad \text{(I.3.1.8)}$$

(P_{η_2} - a.e), and the Radon-Nicodym derivative of $P_{\xi_2|\xi_1}(.|x_1)$ with respect to $P_{\eta_2|\eta_1}(.|x_1)$ is

$$a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) = \frac{a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)}{a_{\xi_1:\eta_1}(x_1)} = \frac{P_{\xi_2|\xi_1}(dx_2|x_1)}{P_{\eta_2|\eta_1}(dx_2|x_1)} \qquad \text{(I.3.1.9)}$$

(P_{η_1} a.e).

Proof. From (I.3.1.6) it follows

$$P_{\xi_1\xi_2}(T) = \int_T a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\,.\,P_{\eta_1\eta_2}(dx_1dx_2) \qquad \text{(I.3.1.10)}$$

so that if $T = T_1 \times X_2$, $T_1 \varepsilon S_1$, from Fubrini's theorem it follows that

$$P_{\xi_1}(T_1) = P_{\xi_1\xi_2}(T_1 \times X_2) = \int_{T_1} [\int_{X_2} a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\, P_{\eta_2|\eta_1}(dx_2|x_1)]\, P_{\eta_1}(dx_1)$$

$$= \int_{T_1} a_{\xi_1:\eta_1}(x_1)\, P_{\eta_1}(dx_1) \qquad (I.3.1.11)$$

where $a_{\xi_1:\eta_1}(x_1)$ is given by (I.3.1.7).

Similarly, if in (I.3.1.10) we take $T = X_1 \times T_2$, $T_2 \varepsilon S_2$, it follows in the same way

$$P_{\xi_2}(T_2) = P_{\xi_1\xi_2}(X_1 \times T_2) = \int_{T_2} [\int_{X_1} a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\, P_{\eta_2|\eta_1}(dx_1|x_2)]\, P_{\eta_2}(dx_2)$$

$$= \int_{T_2} a_{\xi_2:\eta_2}(x_2)\, P_{\eta_2}(dx_2) \qquad (I.3.1.12)$$

where $a_{\xi_2:\eta_2}(x_2)$ is given by (I.3.1.8).

The relation (I.3.1.10) can be written as

$$\int_T P_{\xi_2|\xi_1}(dx_2|x_1)\, P_{\xi_1}(dx_1) = \int_T a_{(\xi_1\ \xi_2):(\eta_1\ \eta_2)}(x_1,x_2)\, P_{\eta_2|\eta_1}(dx_2|x_1) P_{\eta_1}(dx_1)$$

(I.3.1.13)

and using (I.3.1.7) we can write it as

$$\int_T [P_{\xi_2|\xi_1}(dx_2|x_1)\, a_{\xi_1:\eta_1}(x_1) - a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\, P_{\eta_2|\eta_1}(dx_2|x_1)]\, P_{\eta_1}(dx_1) = 0$$

(I.3.1.14)

for any set $T\ \varepsilon\ S_1 \times S_2$, so that

$$P_{\xi_2|\xi_1}(dx_2|x_1)\, a_{\xi_1:\eta_1}(x_1) - a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\, P_{\eta_2|\eta_1}(dx_2|x_1) = 0$$

(I.3.1.15)

(P_{η_1} a.e.). Consequently, denoting

$$a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) = \frac{a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)}{a_{\xi_1:\eta_1}(x_1)} \qquad (I.3.1.16)$$

we obtain

$$P_{\xi_2|\xi_1}(dx_2|x_1) = a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2).P_{\eta_2|\eta_1}(dx_2|x_1) \qquad (I.3.1.17)$$

(P_{η_1} a.e.). From (I.3.1.17) it follows

$$P_{\xi_2|\xi_1}(T_2|x_1) = \int_{T_2} a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)\ P_{\eta_2|\eta_1}(dx_2|x_1) \qquad (I.3.1.18)$$

i.e. $P_{\xi_2|\xi_1}(.|x_1)$ is absolutely continuous with respect to $P_{\eta_2|\eta_1}(.|x_1)$ (P_{η_1} a.e.), with (I.3.1.16) as the Radon-Nicodym derivative.

Let us now denote

$$i_{\xi_1:\eta_1}(x_1) = \log a_{\xi_1:\eta_1}(x_1);\ i_{\xi_2:\eta_2}(x_2) = \log a_{\xi_2:\eta_2}(x_2) \qquad (I.3.1.19)$$

$$i_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) = \log a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) \qquad (I.3.1.20)$$

$$i_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) = \log a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) \qquad (I.3.1.21)$$

<u>Lemma I.6.</u> In each of the relations

$$a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) = a_{\xi_1:\eta_1}(x_1).\ a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)\ (P_{\eta_1}\text{ a.e}) \qquad (I.3.1.22)$$

$$i_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) = i_{\xi_1:\eta_1}(x_1) + i_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)\ (P_{\eta_1}\text{ a.e}) \qquad (I.3.1.23)$$

if two of the quantities are finite, then the third one is also finite and the relation is verified.

In the case that the random vectors ξ_1,η_1 are independent, and the random vectors ξ_2,η_2 are also independent, the relations (I.3.1.22), (I.3.1.23) take the simplified form

$$a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) = a_{\xi_1:\eta_1}(x_1)\ .\ a_{\xi_2:\eta_2}(x_2) \qquad (I.3.1.22')$$

$$i_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) = i_{\xi_1:\eta_1}(x_1) + i_{\xi_2:\eta_2}(x_2) \qquad (I.3.1.23')$$

Proof. The equality (I.3.1.22) follows from (I.3.1.9) and (I.3.1.23) follows from (I.3.1.22), taking into consideration (I.3.1.20), (I.3.1.21). The relations (I.3.1.22'), (I.3.1.23') follow from the relations (I.3.1.22), (I.3.1.23), taking into consideration that in the conditions of independence indicated in the Lemma,

$$a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) = a_{\xi_2:\eta_1}(x_2) \tag{I.3.1.24}$$

$$i_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) = i_{\xi_2:\eta_2}(x_2) \tag{I.3.1.25}$$

Let

$$h[(\xi_2|x_1):(\eta_2|x_1)] = \int_{X_2} a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) \log a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) P_{\eta_2|\eta_1} \cdot (dx_2|x_1) \tag{I.3.1.26}$$

$$h[(\xi_2|\xi_1):(\eta_2|\eta_1)] = \int_{X_1} h[(\xi_2|x_1):(\eta_2|x_1)]\ P_{\xi_1}(dx_1) \tag{I.3.1.27}$$

This is the relative conditional entropy of the ordered pair of random vectors ξ_1,ξ_2 with respect to the ordered pair of random vectors η_1,η_2.

Lemma I.7.

a) If two of the three quantities in the relation

$$h[(\xi_1\xi_2):(\eta_1\eta_2)] = h(\xi_1:\eta_2) + h[(\xi_2|\xi_1):(\eta_2|\eta_1)] \tag{I.3.1.28}$$

are finite, then the third one is also finite and they verify this relation.

b) In the particular case where ξ_1,ξ_2 are independent and η_1,η_2 are independent, the relation (I.3.1.28) takes the simplified form

$$h[(\xi_1\xi_2):(\eta_1\eta_2)] = h(\xi_1:\eta_1) + h(\xi_2:\eta_2) \tag{I.3.1.28'}$$

Proof. Because of (I.3.1.22), we obtain

$$a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) \log a_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2) = [a_{\xi_1:\eta_1}(x_1) \log a_{\xi_1:\eta_1}(x_1)]$$

$$\cdot a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) + [a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) \log a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)]$$

$$\cdot a_{\xi_1:\eta_1}(x_1) \tag{I.3.1.29}$$

and we take the integrals of both members on $P_{\eta_1\eta_2}$ - measure over $X_1 \times X_2$. From (I.3.1.26), (I.3.1.27) it is seen that the integral of the left hand side is $h[(\xi_1\xi_2):(\eta_1\eta_2)]$.

Let us consider now the integral of the first term of the right hand side. We obtain from Fubini's theorem and from (I.3.1.9)

$$\int_{X_1\times X_2} [a_{\xi_1:\eta_1}(x_1) \log a_{\xi_1:\eta_1}(x_1)]\, a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)\, P_{\eta_1\eta_2}(dx_1dx_2) =$$

$$= \int_{X_1\times X_2} [a_{\xi_1:\eta_1}(x_1) \log a_{\xi_1:\eta_1}(x_1)] \cdot \frac{P_{\xi_2|\xi_1}(dx_2|x_1)}{P_{\eta_2|\eta_1}(dx_2|x_1)} \cdot P_{\eta_2|\eta_1}(dx_2|x_1) . P_{\eta_1}(dx_1) =$$

$$= \int_{X_1} a_{\xi_1:\eta_1}(x_1) \log a_{\xi_1:\eta_1}(x_1) . P_{\eta_1}(dx_1) . \int_X P_{\xi_2|\xi_1}(dx_2|x_1) =$$

$$= \int_{X_1} a_{\xi_1:\eta_1}(x_1) \log a_{\xi_1:\eta_1}(x_1) . P_{\eta_1}(dx_1) = h(\xi_1:\eta_1) = \qquad \text{(I.3.1.29')}$$

Let us consider now the integral of the second term of the right hand side. We obtain from Fubini's theorem and (I.3.1.7), taking in consideration (I.3.1.26), (I.3.1.27), that

$$\int_{X_1\times X_2} [a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) \log a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)] a_{\xi_1:\eta_1}(x_1)\; P_{\eta_1\eta_2}(dx_1dx_2) =$$

$$= \int_{X_1\times X_2} [a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) \log a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)]\, \frac{P_{\xi_1}(dx_1)}{P_{\eta_1}(dx_1)} \cdot$$

$$\cdot P_{\eta_2|\eta_1}(dx_2|x_1)\; P_{\eta_1}(dx_1) =$$

$$= \int_{X_1} P_{\xi_1}(dx_1) \int_{X_2} [a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2) \log a_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)]\; P_{\eta_2|\eta_1}(dx_2|x_1) =$$

$$= \int_{X_1} h[(\xi_2|x_1):(\eta_2|x_1)]\; P_{\xi_1}(dx_1) = h[(\xi_2|\xi_1):(\eta_2|\eta_1)] \qquad \text{(I.3.2.29'')}$$

which proves our Lemma I.7 part (a). Part (b) follows from part (a) taking into consideration (I.3.1.25).

I.3.2. Let: (Ω_i,Σ_i) be a measurable space and (Ω_i,Σ_i,P_i) a probability space, where Ω_i is a set of elements ω_i, Σ_i a σ-algebra of subsets of Ω_i, P_i a probability measure on Σ_i $(1 \leq i \leq n)$.

We consider the random vectors ξ_i,η_i, defined on this probability space with values in the measureable space (X_i,S_i) where X_i is a set of elements x_i, S_i a σ-algebra of subsets of X_i $(1 \leq i \leq n)$. Let

$$A \subset \{1,2,\ldots,n\}, \tag{I.3.2.1}$$

and

$$(\Omega_A,\Sigma_A) = \underset{i\varepsilon A}{X} (\Omega_i,\Sigma_i)$$

be a measurable space, where Ω_A is the set of elements $\omega_A = \{\omega_i,\ i \varepsilon A\}$ and $\Sigma_A = \underset{i\varepsilon A}{X} \ \Sigma_i$.

In the particualr case where $A = \{1,2,\ldots k\}$, we denote $\omega_A = \omega^{(k)}$, $\Omega_A = \Omega^{(k)}$, $\Sigma_A = \Sigma^{(k)}$ $(1 \leq k \leq n)$. Let P_A be a probability measure on Σ_A with marginals P_i $(i \varepsilon A)$. So (Ω_A,Σ_A,P_A) is a probability space.

We consider the random vectors $\xi_A(\omega_A)$, $\eta_A(\omega_A)$ given on the probability space $(\Omega_A, \Sigma_A, P_A)$ by

$$\xi_A(\omega_A) = \{\xi_i(\omega_i);\ i \varepsilon A\};\ \eta_A(\omega_A) = \{\eta_i(\omega_i);\ i \varepsilon A\} \tag{I.3.2.2}$$

with values in the measurable space $(X_A,S_A) = \underset{i\varepsilon A}{X} (X_i,S_i)$ where $x_A = \{x_i,\ i \varepsilon A\}$, with $x_i \varepsilon X_i$ $(i \varepsilon A)$, are elements of X_A and $S_A = \underset{i\varepsilon A}{X} S_i$.

If $A = \{1,2,\ldots,k\}$, then we denote $x_A = x^{(k)}$, $X_A = X^{(k)}$, $\xi_A = \xi^{(k)}$, $\eta_A = \eta^{(k)}$.

Let

$$P_{\xi_i}(T_i) = P_i\{\omega_i;\ \xi_i(\omega_i) \varepsilon T_i\},\ T_i \varepsilon S_i \quad (1 \leq i \leq n) \tag{I.3.2.3}$$

$$P_{\eta_i}(T_i) = P_i\{\omega_i;\ \eta_i(\omega_i) \varepsilon T_i\},\ T_i \varepsilon S_i \quad (1 \leq i \leq n) \tag{I.3.2.3'}$$

be the probability measures of the random vectors ξ_i,η_i $(1 \leq i \leq n)$

and let

$$P_{\xi_A}(T_A) = P_A\{\omega_A;\ \xi_A(\omega_A) \in T_A\},\ T_A \in S_A \tag{I.3.2.4}$$

$$P_{\eta_A}(T_A) = P_A\{\omega_A;\ \eta_A(\omega_A) \in T_A,\ T_A \in S_A \tag{I.2.3.4'}$$

<u>Lemma I.8.</u>

If $P_{\xi^{(n)}}$ is absolutely continuous with respect to $P_{\eta^{(n)}}$ and the Radon-Nicodym derivative of $P_{\xi^{(n)}}$ with respect to $P_{\eta^{(n)}}$ is

$$a_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \frac{P_{\xi^{(n)}}(dx^{(n)})}{P_{\eta^{(n)}}(dx^{(n)})}, \tag{I.2.3.5}$$

then we will prove the following:

(a) for any $A \subset \{1,2,\ldots,n\}$, the probability measure P_{ξ_A} is absolutely continuous with respect to P_{η_A} and the Radon-Nicodym derivative of P_{ξ_A} with respect to P_{η_A} is given by the relation

$$a_{\xi_A:\eta_A}(x_A) = \int_{X_{\bar{A}}} a_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}).P_{\eta_{\bar{A}}|\eta_A}(dx_{\bar{A}}|x_A) \tag{I.3.2.6}$$

where $P_{\eta_{\bar{A}}|\eta_A}$ is a conditional marginal measure of $P_{\eta^{(n)}}$, and $\bar{A} = \{1,2,\ldots,n\} - A$.

(b) for any $A_1 \subset \{1,2,\ldots,n\}$, $A_2 \subset \{1,2,\ldots,n\}$, $A_1 \cap A_2 = \emptyset$, the probability measure

$$P_{\xi_{A_2}|\xi_{A_1}}(.|x_{A_1}) \tag{I.3.2.7}$$

(conditional marginal measure of $P_{\xi^{(n)}}$) is absolutely continuous with respect to the probability measure

$$P_{\eta_{A_2}|\eta_{A_1}}(.|x_{A_1}) \tag{I.2.3.7'}$$

(conditional marginal measure of $P_{\eta^{(n)}}$) and the corresponding Radon-

Nicodym derivative is

$$a_{(\xi_{A_2}|\xi_{A_1}):(\eta_{A_2}|\eta_{A_1})}(x_{A_1\cup A_2}) = \frac{a_{\xi_{A_1\cup A_2}:\eta_{A_1\cup A_2}}(x_{A_1\cup A_2})}{a_{\xi_{A_1}:\eta_{A_1}}(x_{A_1})} =$$

$$= \frac{P_{\xi_{A_2}|\xi_{A_1}}(dx_{A_2}|x_{A_1})}{P_{\eta_{A_2}|\eta_{A_1}}(dx_{A_2}|x_{A_1})} \qquad (P_{A_1}\ \text{a.e.}) \qquad \text{(I.3.2.8)}$$

Proof.

(a) Let us consider the result (a) in Lemma I.9 for

$$\xi_1' = \xi_{\bar{A}},\ \xi_2' = \xi_A;\ \eta_1' = \eta_{\bar{A}},\ \eta_2' = \eta_A \qquad \text{(I.3.2.9)}$$

where $A \subset \{1,2,\ldots,n\}$, $\bar{A} = \{1,2,\ldots,n\} - A$. In this case

$$P_{\xi_1'\xi_2'} = P_{\xi(n)},\quad P_{\eta_1'\eta_2'} = P_{\eta(n)} \qquad \text{(I.3.2.10)}$$

and because $P_{\xi(n)}$ is absolutely continuous with respect to $P_{\eta(n)}$, it follows that P_{ξ_A} is absolutely continuous with respect to P_{η_A}. The relation (I.3.2.6) follows from (I.3.1.7).

(b) Let A_1, A_2 be two mutually exclusive subsets of $\{1,2,\ldots,n\}$ so that $A_1 \cup A_2$ is also a subset of the same, so from part (a) above it follows that $P_{\xi_{A_1\cup A_2}}$ is absolutely continuous with respect to $P_{\eta_{A_1\cup A_2}}$.

Let

$$\xi_1' = \xi_{A_1},\ \xi_2' = \xi_{A_2};\ \eta_1' = \eta_{A_1},\ \eta_2' = \eta_{A_2} \qquad \text{(I.3.2.11)}$$

so that

$$P_{\xi_{A_1}\xi_{A_2}} = P_{\xi_{A_1\cup A_2}},\quad P_{\eta_{A_1}\eta_{A_2}} = P_{\eta_{A_1\cup A_2}} \qquad \text{(I.3.2.12)}$$

From Lemma I.9(b) it follows that

$$P_{\xi_{A_2}|\xi_{A_1}}(.|x_{A_1}) \qquad \text{(I.3.2.13)}$$

is absolutely continuous with respect to

$$P_{\eta_{A_2}|\eta_{A_1}}(.|x_{A_1}). \qquad \text{(I.3.2.13')}$$

The relation (I.3.2.8) follows from (I.3.1.9)

The Lemma is proved.

In what follows we will be particularly interested in the case where $A_2 = \{k\}$, $A_1 = \{1,2,\ldots,k-1\}$ $(1 \le k \le n)$ so that $\xi_{A_1} = \xi^{(k-1)}$, $\xi_{A_2} = \xi_k$. From Lemma I.8, it follows that

$$P_{\xi_k|\xi^{(k-1)}}(\cdot|x^{(k-1)}) \tag{I.3.2.14}$$

is absolutely continuous with respect to

$$P_{\eta_k|\eta^{(k-1)}}(\cdot|x^{(k-1)}) \tag{I.3.2.14'}$$

($P_{\eta^{(k-1)}}$ - a.e.) and the corresponding Radon-Nicodym derivative is

$$a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) = \frac{P_{\xi_k|\xi^{(k-1)}}(dx_k|x^{(k-1)})}{P_{\eta_k|\eta^{(k-1)}}(dx_k|x^{(k-1)})} \tag{I.3.2.15}$$

Let

$$i_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \log a_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) \tag{I.3.2.16}$$

$$i_{(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})}(x_{k-1},x_k) = \log a_{(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})}(x_{k-1},x_k) \tag{I.3.2.17}$$

$$i_{\xi_k:\eta_k}(x_k) = \log a_{\xi_k:\eta_k}(x_k) \tag{I.3.2.18}$$

$$i_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) = \log a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) \tag{I.3.2.19}$$

<u>Lemma I.9.</u>

(a) In each of the relations

$$a_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \prod_{k=1}^{n} a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) \tag{I.3.2.20}$$

$$i_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \sum_{k=1}^{n} i_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) \tag{I.2.3.21}$$

if n of the quantities are finite, then the n+1 - th is also finite and the relation is satisfied.

(b) In the case the random vectors ξ_i $(1 \leq i \leq n)$ form a simple Markov chain, and the random vectors η_i $(1 \leq i \leq n)$ form a simple Markov chain, the relations (I.3.2.20), (I.3.2.21) take the simplified form

$$a_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \prod_{k=1}^{n} a_{(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})}(x_{k-1},x_k) \tag{I.3.2.20'}$$

$$i_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \sum_{k=1}^{n} i_{(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})}(x_{k-1},x_k) \tag{I.3.2.21'}$$

(c) In the case the random vectors $\xi_i (1 \leq i \leq n)$ are independent in their totality and the random vectors η_i $(1 \leq i \leq n)$ are independent in their totality, the relations (I.3.2.20), (I.3.2.21) take the simplified form

$$a_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}) = \prod_{k=1}^{n} a_{\xi_k:\eta_k}(x_k) \tag{I.3.2.20''}$$

$$i_{\xi(n):\eta(n)}(x^{(n)}) = \sum_{k=1}^{n} i_{\xi_k:\eta_k}(x_k) \tag{I.3.2.21''}$$

<u>Proof</u>.

(a) From the identity

$$\frac{P_{\xi^{(n)}}(dx^{(n)})}{P_{\eta^{(n)}}(dx^{(n)})} = \prod_{k=1}^{n} \frac{P_{\xi_k|\xi^{(k-1)}}(dx_k|x^{(k-1)})}{P_{\eta_k|\eta^{(k-1)}}(dx_k|x^{(k-1)})} \tag{I.3.2.22}$$

taking into consideration (I.3.2.5), (I.3.2.15) follows (I.3.2.20). From (I.3.2.20), taking into consideration (I.3.2.16), (I.3.2.19) follows (I.3.2.21).

(b) In the conditions of markovian dependence indicated in the Lemma,

$$a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) = a_{(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})}(x_{k-1},x_k) \tag{I.3.2.23}$$

$$i_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) = i_{(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})}(x_{k-1},x_k) \tag{I.3.2.24}$$

Consequently the relations (I.3.2.20), (I.3.2.21) take the form (I.3.2.20'), (I.3.2.21').

(c) In the conditions of independence indicated in the Lemma

$$a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) = a_{\xi_k:\eta_k}(x_k) \tag{I.3.2.23'}$$

$$i_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) = i_{\xi_k:\eta_k}(x_k) \tag{I.3.2.24'}$$

Consequently the relations (I.3.2.20), (I.3.2.21) take the form (I.3.2.20''), (I.3.2.21'').

The Lemma is proved.

Let

$$h[(\xi_k|x^{(k-1)}):(\eta_k|x^{(k-1)})]$$
$$= \int_{X_k} a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) \log a_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}) \cdot$$
$$\cdot P_{\eta_k|\eta^{(k-1)}}(dx_k|x^{(k-1)}) \tag{I.3.2.25}$$

$$h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})]$$
$$= \int_{X^{(k-1)}} h[(\xi_k|x^{(k-1)}):(\eta_k|x^{(k-1)})] \, P_{\xi^{(k-1)}}(dx^{(k-1)}) \tag{I.3.2.25'}$$

<u>Lemma I.10</u>.

(a) If n of the n + 1 quantities in the relation

$$h(\xi^{(n)}:\eta^{(n)}) = \sum_{k=1}^{n} h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})] \tag{I.3.2.26}$$

are finite, then the n + 1 - th one is finite also and the relation is satisfied.

(b) In the case the random vectors ξ_i $(1 \leq i \leq n)$ form a simple Markov chain, and the random vectors η_i $(1 \leq i \leq n)$ form a simple Markov chain, the relation (I.3.2.26) takes the simplified form

$$h(\xi^{(n)}:\eta^{(n)}) = \sum_{k=1}^{n} h[(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})] \tag{I.3.2.26'}$$

(c) In the case the random vectors ξ_i $(1 \leq i \leq n)$ are independent in their totality, and the random vectors η_i $(1 \leq i \leq n)$ are independent in their totality, the relation (I.3.2.26) takes the simplified form

$$h(\xi^{(n)}:\eta^{(n)}) = \sum_{k=1}^{n} h(\xi_k:\eta_k) \qquad \text{(I.3.2.26'')}$$

Here we denoted

$$h[(\xi_1|\xi^{(0)}):(\eta_1|\eta^{(0)})] = h(\xi_1:\eta_1) \qquad \text{(I.3.2.27)}$$

$$h[(\xi_1|\xi_0):(\eta_1|\eta_0)] = h(\xi_1:\eta_1) \qquad \text{(I.3.2.27')}$$

Proof. From Lemma I.12(b) with $A_1 = \{1,2,\dots,k\}$, $A_2 = B_k = \{k,k+1,\dots,n\}$, it follows that if $P_{\xi^{(n)}}$ is absolutely continuous with respect to $P_{\eta^{(n)}}$ then,

$$P_{\xi_{B_k}|\xi^{(k-1)}}(.|x^{(k-1)}) \qquad \text{(I.3.2.28)}$$

is absolutely continuous with respect to

$$P_{\eta_{B_k}|\eta^{(k-1)}}(.|x^{(k-1)}) \qquad \text{(I.3.2.28')}$$

for $(k = 2,3,\dots,n)$. Let now

$$\xi_1' = \xi_k;\ \xi_2' = \xi_{B_{k+1}};\ \eta_1' = \eta_k;\ \eta_2' = \eta_{B_{k+1}};\ B_k = \{k,\dots,n\} \qquad \text{(I.3.2.29)}$$

So

$$\xi_1'\xi_2' = \xi_{B_k},\ \eta_1'\eta_2' = \eta_{B_k} \qquad \text{(I.3.2.30)}$$

and consequently from (I.3.1.28) it follows

$$h[(\xi_{B_k}|\xi^{(k-1)}):(\eta_{B_k}|\eta^{(k-1)})] = h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})+$$

$$+ h[(\xi_{B_{k+1}}|\xi^{(k)}):(\eta_{B_{k+1}}|\eta^{(k)})] \quad (1 \leq k \leq n-1) \qquad \text{(I.3.2.31)}$$

so that

$$\sum_{k=1}^{n-1} h[(\xi_{B_k}|\xi^{(k-1)}):(\eta_{B_k}|\eta^{(k-1)})] = \sum_{k=1}^{n-1} h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})]+$$

$$+ \sum_{k=1}^{n-1} h[(\xi_{B_{k+1}}|\xi^{(k)}):(\eta_{B_{k+1}}|\eta^{(k)})] \qquad \text{(I.3.2.32)}$$

where

$$\xi_{B_1} = \xi^{(n)} \ , \ \eta_{B_1} = \eta^{(n)} \qquad \text{(I.3.2.33)}$$

From (I.3.2.32) follows (I.3.2.26).

(b) In the conditions of markovian dependence indicated in the Lemma, the relations (I.3.2.23), (I.3.2.24)are satisfied and this implies

$$h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})] = h[(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1})] \qquad \text{(I.3.2.34)}$$

so that (I.3.2.26) takes the form (I.3.2.26').

(c) In the condition,independence indicated in the Lemma of the relations (I.3.2.23'), (I.3.2.24') are satisfied and this implies

$$h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})] = H(\xi_k:\eta_k) \qquad \text{(I.3.2.35)}$$

so that (I.3.2.26) takes the form (I.3.2.26'').

I.4. An extension

I.4.1. Let ξ, η, ζ, be three random vectors, with the same values x_i, and let

$$P_\xi(x_i) = P(\xi = x_i);\ P_\eta(x_i) = P(\eta = x_i);\ P_\zeta(x_i) = P(\zeta = x_i)\ \ (1 \le i \le n) \tag{I.4.1.1}$$

In analogy to (I.1.1.2), the relative entropy of ξ with respect to η from the point of view of ζ (or of P_ξ with respect to P_η from the point of view of P_ζ)is given by the expression

$$h(\xi : \eta ; \zeta) = h(P_\xi : P_\eta ; P_\zeta) = \sum_{i=1}^{n} P_\zeta(x_i) \log \frac{P_\xi(x_i)}{P_\eta(x_i)} \tag{I.4.1.2}$$

where for $a \geq 0$ we consider $0 \log \frac{0}{a} = 0$.

I.4.2 Now let (Ω, Σ, P) be a probability space, where Ω is a set of elements ω, Σ a σ-algebra of subsets of Ω, P a probability measure on Σ. We consider three random vectors ξ, η, ζ, defined on this probability space, with values in the measure space (X, S, μ), where X is a set of elements x, S a σ-algebra of subsets of X, μ - a measure on S.

Let

$$P_\xi(T) = P\{\omega;\ \xi(\omega) \in T\};\ P_\eta(T) = P\{\omega;\ \eta(\omega) \in T\};\ P_\zeta(T) = P\{\omega;\ \zeta(\omega) \in T\},\ T \in S \tag{I.4.2.1}$$

be their probability measures.

With $a_{\xi:\eta}(x)$ defined in (I.1.8.2) and $i_{\xi:\eta}(x)$ defined in (I.1.8.4), in analogy to (I.1.8.3), we define the quantity

$$h(\xi : \eta ; \zeta) = \int_X i_{\xi:\eta}(x) . P_\zeta(dx) \tag{I.4.2.2}$$

as the relative entropy of ξ with respect to η from the point of view of ζ.

In the particular case that the probability measures P_ξ, P_η, P_ζ are defined in terms of densities $\pi_\xi(x)$, $\pi_\eta(x)$, $\pi_\zeta(x)$, with respect to a measure μ, the integral formula (I.4.2.2) reduces to

$$h(\xi : \eta ; \zeta) = \int_X \pi_\zeta(x) \log \frac{\pi_\xi(x)}{\pi_\eta(x)}\, dx \tag{I.4.2.3}$$

where the integration is on μ - measure. Obviously, if $\zeta = \xi$, the expression (I.4.1.2) reduces (I.1.1.2), the expression (I.4.2.2) reduces to (I.1.8.3) and (I.4.2.3) reduces to (I.1.8.5), i.e.

$$h(\xi:\eta;\xi) = h(\xi:\eta) \tag{I.4.2.4}$$

In analogy with (I.3.1.26) we define the quantity

$$h[(\xi_2|x_1):(\eta_2|x_1);(\zeta_2|x_1)] = \int_{X_2} i_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)\, P_{\zeta_2|\zeta_1}(dx_2|x_1) \tag{I.4.2.5}$$

and in analogy with (I.3.1.27) we define the quantity

$$h[(\xi_2|\xi_1):(\eta_2|\eta_1);(\zeta_2|\zeta_1)] = \int_{X_1} h[(\xi_2|x_1):(\eta_2|x_1):(\zeta_2|x_1)]\, P_{\zeta_1}(dx_1) \tag{I.4.2.6}$$

This quantity is the relative conditional entropy of the ordered pair of random vectors ξ_1,ξ_2 with respect to the ordered pair of random vectors η_1,η_2 from the point of view of ζ_1,ζ_2.

Lemma I.11.

(a) If two of the three quantities in the relation

$$h[(\xi_1\xi_2):(\eta_1\eta_2);(\zeta_1\zeta_2)] = h(\xi_1:\eta_1;\zeta_1) + h[(\xi_2|\xi_1):(\eta_2|\eta_1);(\zeta_2|\zeta_1)] \tag{I.4.2.7}$$

are finite, then the third one is also finite and they verify this relation.

(b) In the particular case where ξ_1,ξ_2 are independent, and η_1,η_2 are independent, the relation (I.4.2.7) takes the simplified form

$$h[(\xi_1\xi_2):(\eta_1\eta_2);(\zeta_1\zeta_2)] = h(\xi_1:\eta_1;\zeta_1) + h(\xi_2:\eta_2;\zeta_2) \tag{I.4.2.7'}$$

Proof. Integrating both sides of relation (I.3.1.23) on $P_{\zeta_1\zeta_2}(dx_1,dx_2)$

it follows from Fubini's theorem that

$$h[(\xi_1\xi_2):(\eta_1\eta_2);(\zeta_1\zeta_2)] = \int_{X_1\times X_2} i_{(\xi_1\xi_2):(\eta_1\eta_2)}(x_1,x_2)\, P_{\zeta_1\zeta_2}(dx_1,dx_2) =$$

$$= \int_{X_1} i_{\xi_1:\eta_1}(x_1)\, P_{\zeta_1}(dx_1) \int_{X_2} P_{\zeta_2|\zeta_1}(dx_2|x_1) +$$

$$+ \int_{X_1} P_{\zeta_1}(dx_1) \int_{X_2} i_{(\xi_2|\xi_1):(\eta_2|\eta_1)}(x_1,x_2)\, P_{\zeta_2|\zeta_1}(dx_2|x_1) \qquad (I.4.2.8)$$

Taking into consideration (I.4.2.2), (I.4.2.5), (I.4.2.6) we obtain (I.4.2.7). Taking into consideration (I.3.1.25), from (I.4.2.7) follows (I.4.2.7').

I.4.3. In what follows, we will use the concepts and notations in I.3.2. Similar to the random vectors ξ_i, η_i, we consider the random vectors ζ_i, defined on the same space $(\Omega_i, \Sigma_i, P_i)$ and with values in the same measurable space (X_i, S_i) $(1 \leq i \leq n)$. Similar to the random vectors $\xi_A(\omega_A)$, $\eta_A(\omega_A)$, we consider the random vector

$$\zeta_A(\omega_A) = \{\zeta_i(\omega_i),\ i \in A\} \qquad (I.4.3.1)$$

defined on the same space $(\Omega_A, \Sigma_A, P_A)$ and with the values in the same measurable space (X_A, S_A).

If $A = \{1,2,\ldots,k\}$, let $\zeta_A = \zeta^{(k)}$.

Similar to $P_{\xi_i}(T_i)$, $P_{\eta_i}(T_i)$ in (I.3.2.3), (I.3.2.3'), we define

$$P_{\zeta_i}(T_i) = P_i\{\omega_i;\ \zeta_i(\omega_i) \in T_i\},\ T_i \in S_i,\ (1 \leq i \leq n) \qquad (I.4.3.2)$$

as the probability measure of the random vector ζ_i $(1 \leq i \leq n)$ and similar to $P_{\xi_A}(T_A)$, $P_{\eta_A}(T_A)$ in (I.3.2.4), (I.3.2.4') we define

$$P_{\zeta_A}(T_i) = P_A\{\omega_A;\ \zeta_A(\omega_A) \in T_A\},\ T_A \in S_A \qquad (I.4.3.3)$$

Similar to (I.3.2.14), (I.3.2.14'), the measure

$$P_{\zeta_k|\zeta^{(k-1)}}(\cdot|x^{(k-1)}) \tag{I.4.3.4}$$

is a conditional marginal measure from $P_{\zeta^{(n)}}$.

Let

$h[(\xi_k|x^{(k-1)}):(\eta_k|x^{(k-1)});(\zeta_k|x^{(k-1)}) =$

$h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)});(\zeta_k|\zeta^{(k-1)})] =$

$$= \int_{X^{(k-1)}} h[(\xi_k|x^{(k-1)}):(\eta_k|x^{(k-1)});(\zeta_k|x^{(k-1)})]\ P_{\zeta^{(k-1)}}(dx^{(k-1)}) \tag{I.4.3.5}$$

Lemma I.12.

(a) If n of the n + 1 quantities in the relation

$$h(\xi^{(n)}:\eta^{(n)};\zeta^{(n)}) = \sum_{k=1}^{n} h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)});(\zeta_k|\zeta^{(k-1)})] \tag{I.4.3.6}$$

are finite, then the $n+1$ -th one is finite also and the relation is satisfied.

(b) In the case the random vectors ξ_i $(1 \leq i \leq n)$ form a simple Markov chain, and the random vectors η_i $(1 \leq i \leq n)$ form a simple Markov chain, the relation (I.4.3.6) takes the simplified form

$$h(\xi^{(n)}:\eta^{(n)};\zeta^{(n)}) = \sum_{k=1}^{n} h[(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1});(\zeta_k|\zeta_{k-1})] \tag{I.4.3.6'}$$

(c) In the case the random vectors $\xi_i (1 \leq i \leq n)$ are independent in their totality, and the random vectors η_i $(1 \leq i \leq n)$ are also independent in their totality, the relation (I.4.3.6) takes the simplified form

$$h(\xi^{(n)}:\eta^{(n)};\zeta^{(n)}) = \sum_{k=1}^{n} h(\xi_k:\eta_k;\zeta_k) \tag{I.4.3.6''}$$

Here we denoted $h[(\xi_1|\xi^{(0)}):(\eta_1|\eta^{(0)});(\zeta_1|\zeta^{(0)})] = h[(\xi_1|\xi_0):(\eta_1|\eta_0);(\zeta_1|\zeta_0)] =$ $= h(\xi_1:\eta_1;\zeta_1)$.

Proof.

(a) Integrating both sides of (I.3.2.21) on $P_{\zeta^{(n)}}(dx^{(n)})$, it follows from Fubini's theorem that if $B_k = \{k+1,\ldots,n\}$

$$h[\xi^{(n)}:\eta^{(n)};\zeta^{(n)}] = \int_{X^{(n)}} i_{\xi^{(n)}:\eta^{(n)}}(x^{(n)}).P_{\zeta^{(n)}}(dx^{(n)}) =$$

$$= \sum_{k=1}^{n} \int_{X^{(k)}} i_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)})\ P_{\zeta^{(k)}}(dx^{(k)}) \int_{X_{B_k}} P_{\zeta_{B_k}}(dx_{B_k}|\ x^{(k)}) =$$

$$= \sum_{k=1}^{n} \int_{X^{(k)}} i_{(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)})}(x^{(k)}).P_{\zeta^{(k)}}(dx^{(k)}) =$$

$$= \sum_{k=1}^{n} h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)});(\zeta_k|\zeta^{(k-1)})] \qquad \text{(I.4.3.7)}$$

(b) In the conditions of markovian dependence indicated in the Lemma, the relations (I.3.2.23), (I.3.2.24) are satisfied and this implies

$$h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)});(\zeta_k|\zeta^{(k-1)})] = h[(\xi_k|\xi_{k-1}):(\eta_k|\eta_{k-1});(\zeta_k|\zeta_{k-1})], \qquad \text{(I.4.3.8)}$$

so that (I.4.3.6) *takes the form* (I.4.3.6').

(c) In the conditions of independence indicated in the Lemma, the relations (I.3.2.23'), (I.3.2.24') are satisfied and this implies

$$h[(\xi_k|\xi^{(k-1)}):(\eta_k|\eta^{(k-1)});(\zeta_k|\zeta^{(k-1)})] = h(\xi_k:\eta_k;\ \zeta_k) \qquad \text{(I.4.3.9)}$$

so that (I.4.3.6) takes the form (I.4.3.6'').

I.5. Comments

Lemma I.1 - is a known result (See [9] page 17, Ex. 3.2) proved here with typical elementary means. The Author has not seen this proof elsewhere. In [9] it is proved in a complicated way and in [1] in a most complicated way.

Lemma I.2 - belongs to the Author; the proof uses the method used in [3], page 203 in a particular case.

Lemma I.3 - In the case $r = 2$, part a) is proved in [1]. For $r > 2$ part a) and part b) together with the proofs belong to the Author.

Theorem I.3 - is presented in [11] without proof, references being made to a particular case in [1]. Our proof follows A. Feinstein's remarks to Ch. 2 in [11], with many additions and clarifications.

Theorem I.4 - is presented in [11] without proof. The proof given here belongs to the author.

Theorem I.4' - belongs to the Author.

Lemma I.5 - belongs to the Author.

Lemma I.6 - belongs to the Author.

Lemma I.7 - belongs to the Author.

Theorem I.7 - is presented in [11] without proof; this proof belongs to the Author.

Theorem I.6 - belongs to the Author.

Theorem I.7 - can be found in [11], but the proof has been partially changed for the sake of clarity.

Theorem I.8 - Part a) can be found in [8]. While following the proof of a) in [8], the Author changed the order of presentation for the sake of clarity. Part b) represents an amelioration of a statement in [8], and it belongs to the Author, even that methods from [8] are used.

Theorem I.9 - while following the proof in [8], the Author changed the order of presentation for the sake of clarity.

Theorem I.10 - follows [8], with some clarifications in the proof.

Lemma I.5 - belongs to the Author.

Lemma I.6 - belongs to the Author.

Lemma I.7 - belongs to the Author. See also [12].

Lemma I.8 - belongs to the Author.

Lemma I.9 - belongs to the Author.

Lemma I.10 - belongs to the Author.

Lemma I.11 - belongs to the Author. See also [15].

Lemma I.12 - belongs to the Author.

I.6. References

[1] Dobrushin, R. L. General formulation of Shannon's basic theorems of the theory of information. Usp. Mat. Natk. Vol. 14, N. 6, 1959, p. 3-104. (Russian) English translations in "Mathematical Statistics and Probability" Published by the Institute of Math. Statistics and the Amer. Math Soc. 1961, v. 1, p. 323-436.

[2] Gelfand, I. M., Kolmogorov, A.N. and Yaglom, A. M. On the general definition of the quantity of information. SSSR, v. 111, N. 4, p. 745-748, 1956 (Russian).

[3] Gelfand, I. M. and Yaglom, A. M. Calculation of the amount of information about a random function contained in another such function. Usp. Mat. Nauk, v. 12, N. 1, p. 3-52, 1957 (Russian). English translation in "American Mathematical Society Translations", Providence, RI, Series 2, Vol. 12, p. 199-246; 1959.

[4] Halmos, P. R. Measure theory, Van Nostrand Co., Inc. Princeton, NJ, 1950.

[5] Jeffreys, H. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London, Series A, Vol. 186, 1946, p. 453-461.

[6] Jeffreys, H. Theory of probability, Second edition, Oxford, Clarendon Press, 1948.

[7] Kallianpur, G. On the amount of information contained in a σ-field. In the volume "Contributions to Probability and Statistics, Essays in honor of Harold Hotelling", Edited by Ingram Olkin and others. Stanford University Press, 1960, p. 265-273.

[8] Kemperman, J. H. B. On the optimum rate of transmitting information, Annals of Math. Statist., v. 40, N. 6, 1969, p. 2156-2177.

[9] Kullback, S. Information theory and statistics, John Wiley and Sons, Inc., New York, 1959.

[10] Kullback, S. and Leibler, R. A. On information and sufficiency. Annals of Math. Statist. v. 22, p. 72-86, 1951.

[11] Pinsker, M. S. Information and information stability of random variables and processes. Holden Day, Inc., San Francisco, 1964. Russian original 1960.

[12] Rosenblatt-Roth, M. Likelihood ratio and Kullback's divergence in random processes. Proceedings of the 1979 Conference on Information Sciences and Systems, Department of Electrical Engineering, The John's Hopkins University, p. 43-45.

[13] Rosenblatt-Roth, M. Likelihood ratio and its stability in random processes. Proceedings of the 1980 Conference on Information Sciences and Systems, Department of Electrical Engineering and Computer Science, Princeton University, p. 423-426.

[14] Rosenblatt-Roth, M. On the best approximation of random sources of information with Markov chains. Proceedings of the Seventeenth Annual Conference on Information Sciences and Systems, Department of Electrical Engineering and Computer Science, The Johns Hopkins University, p. 255-260, 1983.

[15] Rosenblatt-Roth, M. On some new quantities of information determined by an ordered sequence of random variables and their use in the finding of the best approximation of random processes with Markov chains. Paper presented at the Sixth International Symposium on Multivariate Analysis, July 25-29, 1983. Center for Multivariate Analysis, University of Pittsburgh.

[16] Shannon, C. E. A mathematical theory of communication. Bell System Technical Journal, v. 27, p. 379-423; 623-656, 1948.

[17] Shannon, C. E. and Weaver, W. The mathematical theory of communication. University of Illinois Press, Urbana, IL, 1949.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 86-0036 ~~83-35~~	2. GOVT ACCESSION NO. AD-A164	3. RECIPIENT'S CATALOG NUMBER 093
4. TITLE (and Subtitle) "The relative entropy of a random vector with respect to another random vector"		5. TYPE OF REPORT & PERIOD COVERED Technical - October 1985
		6. PERFORMING ORG. REPORT NUMBER 85-35
7. AUTHOR(s) M. Rosenblatt-Roth		8. CONTRACT OR GRANT NUMBER(s) ~~F49620-85-C-0008~~ F49620-82-K-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis, 515 Thackeray Hall, University of Pittsburgh, Pittsburgh, PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research, Department of the Air Force, Bolling Air Force Base, DC 20332		12. REPORT DATE October 1985
		13. NUMBER OF PAGES 70
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Additivity Theorems; Entropy, and Information Theory.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper, the author discusses some problems connected with the concept of entropy. Part of the paper is expository in nature and the remaining material deals with new results obtained by the author.

DD FORM 1473 1 JAN 73

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

3-86

DTIC